

Rapport de stage Infoscience

Enrichir la base des journaux et interface de déduplication

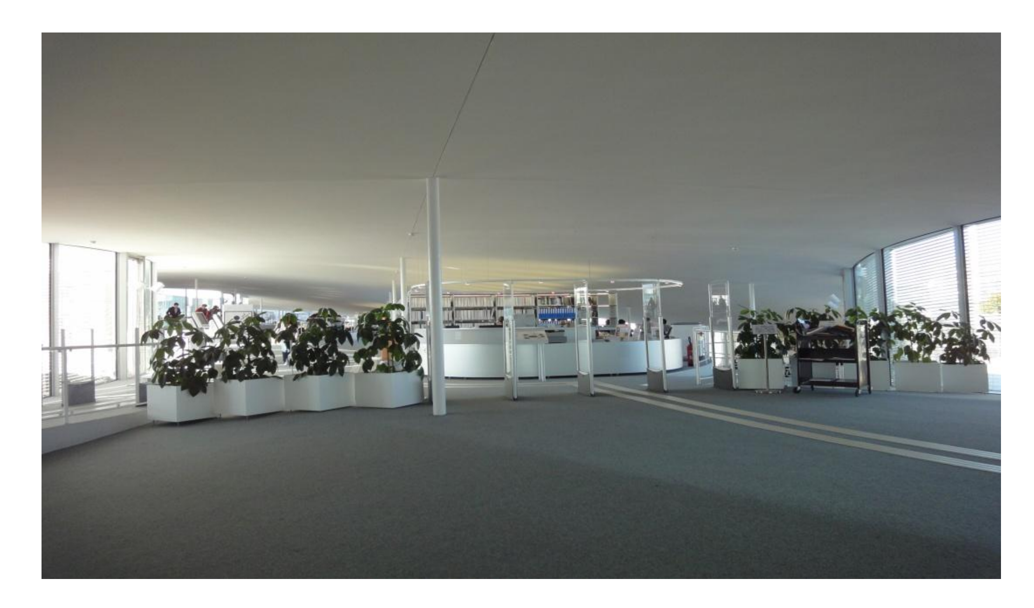


13/11/2010
EPFL-SISB/KIS
Laurence Denoréaz Buclin – HEG-Genève

1 Table des matières

2 Remerciements

Je tiens à remercier toutes les personnes m'ayant permis d'organiser et d'effectuer ma période de stage professionnel. En premier lieu, Pierre Mellier, mon supérieur hiérarchique, qui par son ouverture d'esprit, m'a offert la possibilité de réaliser un travail qui apporterait un plus au groupe DIT-KIS et à la Bibliothèque de l'EPFL. Ensuite David Aymonin, le directeur de la Bibliothèque, pour avoir accepté favorablement ma demande de stage et m'avoir accueilli sans réserve dans son service. Merci à Grégory Favre et à Lionel Walter, qui m'ont conseillé et apporté leur soutien tout au long de ces huit semaines, aux bibliothécaires scientifiques pour leur disponibilité, ainsi qu'à tout le personnel de la Bibliothèque pour son accueil agréable. Je remercie aussi mon fils, Matthieu, pour avoir joué le rôle du photographe et m'avoir permis d'utiliser ses images. Et finalement, merci à mon mari, Bertrand, et à ma collègue Patricia Plaza Gruber pour leur relecture et leurs corrections.



Le bureau d'accueil de la Bibliothèque



Une zone de travail

3 Introduction

3.1 Contexte :

A la fin de la deuxième année des études de spécialiste en information documentaire, un stage professionnel de huit semaines à plein temps doit être effectué. J'ai choisi de faire ce stage à la Bibliothèque du Rolex Learning Center de l'EPFL, sous la direction de M. David Aymonin et en contact direct pour le travail à accomplir avec Grégory Favre, le responsable technique d'Infoscience et Lionel Walter, l'un des bibliothécaires scientifiques du SISB (Service d'Information Scientifique et Bibliothèques).

3.2 Environnement de travail :

Pour effectuer ce stage, une place de travail m'a été octroyée dans le local des informaticiens de la Bibliothèque. Un bureau et un PC ont été mis à ma disposition pour le temps du stage. J'ai eu à disposition toutes les ressources disponibles sur le réseau de l'EPFL, ainsi que toute l'infrastructure du Rolex Learning Center (banque, librairie, restaurant et cafétéria,…).

3.3 Description du stage :

Les tâches à effectuer dans le cadre de ce stage doivent participer au processus d'amélioration qualitative du dépôt d'archives institutionnelles de l'EPFL (Infoscience). En particulier, il s'agit de participer à l'enrichissement de la base des journaux en ajoutant des informations concernant le facteur d'impact, les noms abrégés de certains journaux et des informations sur le droit d'auteur. Et dans un deuxième temps, de fournir des solutions pour la mise en place d'une interface de déduplication des doublons existants dans le dépôt.

3.4 Description d'Infoscience

Infoscience (http://infoscience.epfl.ch/) est le portail d'information scientifique de l'EPFL et le dépôt des archives institutionnelles (repository) des chercheurs de cette institution. Il fournit une interface permettant de signaler et d'enregistrer des publications de toutes sortes (articles de journaux et de conférence, rapports techniques et de travail, monographies, thèses, actes de congrès, présentations, normes et brevets, cours et matériel pédagogique,…).

Ce portail sert de boussole scientifique pour orienter vers les bonnes compétences au sein de l'EPFL. Il propose également une optique de travail respectant le droit d'auteur et promouvant un accès ouvert à la science. Il intègre des outils facilitant la collaboration et le partage. Dans le futur, il fournira des informations analytiques et statistiques sur les données stockées dans le dépôt.

Aujourd'hui, plus de 60'000 références bibliographiques ont été archivées par les collaborateurs de l'EPFL (chercheurs, collaborateurs, bibliothécaires,...). Et parmi celles-ci, plus de 20'000 documents sont disponibles en accès libre.

4 Méthodologie

Le produit résultat demandé dans le cadre de ce stage devra se présenter sous la forme d'un rapport complet incluant un tableau descriptif des outils à évaluer (liste disponible dans le cahier des charges), d'une analyse et synthèse de ce tableau, d'un glossaire de termes spécifiques à la problématique concernée et d'une description des fonctionnalités nécessaires à la mise en place d'une interface de déduplication. Divers autres documents ont été utilisés ou produits pour permettre de mener à bien ce travail et de compléter les informations recueillies (voir la partie Annexes).

Après avoir lu le cahier des charges et obtenu des informations complémentaires durant la première matinée de présence à la Bibliothèque, j'ai préparé différents fichiers (journal de bord, calendrier de travail, liste de liens,...) pour m'aider à mieux organiser mon travail.

Pour bien rester dans la ligne des objectifs du mandat, j'ai fait régulièrement le point avec les deux personnes chargées de superviser mon travail, Lionel et Grégory.

Le travail étant scindé en deux parties, je me suis d'abord préoccupée de l'enrichissement de la base des journaux. Dans un deuxième temps, j'ai traité la « déduplication » en rencontrant les bibliothécaires scientifiques.

Pour la partie enrichissement, j'ai commencé par préparer un glossaire expliquant les termes spécifiques à cette problématique. Ce fichier sera utilisé pour informer les nouveaux collaborateurs de la Bibliothèque au sujet de ces concepts techniques et métier. Ensuite, j'ai analysé de manière détaillée chacun des outils mentionnés dans le cahier des charges et vérifié s'ils présentaient des fonctionnalités intéressantes pour l'enrichissement de la base des journaux d'Infoscience. J'ai également extrait un échantillon de journaux de cette base pour vérifier si on les retrouvait dans ces différents outils.

J'ai eu contact avec Mme Bosia Diabang, la responsable des périodiques électroniques à l'EPFL pour voir s'il y avait lieu de compléter la liste fournie et obtenir des informations supplémentaires à ce sujet. Elle m'a conseillé d'ajouter la base Scopus, ce qui a été un bon complément aux outils que je devais analyser.

J'ai également fait le tour des ressources professionnelles (banques de données et liens divers) enseignées à la HEG pour voir si je pouvais trouver des informations complémentaires au sujet.

J'ai aussi pris contact par email avec d'autres archives institutionnelles à travers le monde pour voir de quelle manière ils procèdent et savoir si il serait possible de partager certaines informations. J'ai analysé leurs réponses et j'ai fait une synthèse de celles-ci.

Pour la partie déduplication, je me suis basée sur le site web d'Infoscience (http://infoscience.epfl.ch), j'ai lu le cahier des charges de la plateforme de gestion bibliographique de l'EPFL (http://infoscience-wiki.epfl.ch/Curator?action=AttachFile&do=get&target=curator-CdC-v1.6.pdf) et j'ai consulté le wiki de travail et d'informations (http://infoscience-wiki.epfl.ch/). Puis, pour définir les fonctionnalités des interfaces à proposer, j'ai recueilli les suggestions des bibliothécaires scientifiques lors d'entretiens individuels.

Après quatre semaines de stage, j'ai présenté un bilan intermédiaire du travail déjà accompli, ainsi qu'une planification des tâches devant encore être effectuées.

Les résultats obtenus et le bilan final seront présentés à la prochaine séance plénière de la Bibliothèque, fixée au 11 octobre 2010.

Une partie des fichiers organisationnels pour ce travail sont disponibles, soit directement dans le texte, soit dans la partie 11 du présent document sous « Annexes ».

5 Définitions

Pour bien cerner les tâches du cahier des charges à accomplir et pouvoir y répondre, je me suis familiarisée, sur les conseils de Lionel, avec un certain nombre de concepts en relation étroite avec le domaine des journaux scientifiques. J'ai recherché les définitions des termes relatifs sur le web, puis j'ai constitué un glossaire avec les différentes informations collectées. On retrouve ces définitions ci-dessous. Un tableau regroupant les définitions détaillées se trouve dans la partie Annexes sous 9.2. Pour le tableau, j'ai choisi de fournir les références des informations trouvées, ainsi que tout ou partie de la définition, plutôt que de faire de la paraphrase.

5.1 La bibliométrie (infométrie ou encore scientométrie)

La bibliométrie est une science qui permet l'analyse quantitative et statistique de documents publiés et de leurs références par rapport à un domaine ou à un chercheur. Les résultats, disponibles sous forme graphique ou autre, donnent la possibilité d'évaluer une activité de recherche, son évolution, son impact, ainsi que l'éventuelle émergence de nouveaux thèmes techniques.

La discipline se met en place dès le début du XXe siècle *(Lotka, Bradford ou Zipf),* mais elle acquiert ses lettres de noblesse à partir des années 1950, grâce à *Derek John de Solla Price.* Le point commun entre ces différentes personnes est leur formation axée sur les disciplines : mathématique, statistique ou linguistique.

5.2 Le facteur d'impact

Le facteur d'impact permet de mesurer l'importance d'une revue scientifique. A l'origine de ce *système bibliographique de mesure de la littérature scientifique*, on trouve Eugene Garfield. Ce dernier a mis en place en 1955 les moyens permettant de suivre de manière fiable et rapide les recherches scientifiques et d'éviter la fraude, les données incomplètes ou dépassées de certaines avancées technologiques. Cela permet d'exercer un contrôle sur la littérature scientifique. Les résultats du facteur d'impact sont disponibles dans le "Journal Citation Reports" de Thomson Reuters.

Ce moyen d'évaluation est aussi utilisé par les sponsors et autres bailleurs de fonds pour le financement des recherches scientifiques.

Le facteur d'impact est souvent décrié comme méthode d'analyse et de valorisation de l'information scientifique. En effet, il permet plutôt de connaître le degré de visibilité d'une publication, que ses qualités. Il faut donc rester prudent avec ces données et les combiner avec d'autres mesures.
Une des manières de contourner le facteur d'impact et d'avoir une certaine fiabilité est d'utiliser d'autres méthodes de mesures.

5.3 Comment se calcule le facteur d'impact

Le facteur d'impact se calcule sur une certaine durée (en général trois ans).
Il s'articule sur le nombre moyen de fois qu'une revue est citée durant les deux années après sa date de publication.
Prenons comme exemple, le facteur d'impact 2009 pour un périodique:

A est le nombre de fois que des articles publiés en 2007 et 2008 sont cités dans la totalité des journaux indexés pendant l'année 2009.

B est le nombre d'articles et autres papiers publiés en 2007 et 2008 dans la revue concernée.

Le facteur d'impact de cette revue pour 2009 est le résultat d'A/B.

5.4 Les alternatives au facteur d'impact

Pour pallier à certains manques du facteur d'impact, des chercheurs ont conçu d'autres méthodes d'évaluation des articles scientifiques, comme par exemple le h-index ou h-index de J E. Hirsch (http://adsabs.harvard.edu/abs/2005physics...8025H) ou encore le e-index de Chun-Ting Zhang (http://www.plosone.org/article/info:doi/10.1371/journal.pone.0005429)[1].
On retrouve certains de ces index dans les bases de données des publications. Scopus, par exemple, référence le h-index.

Ces différents systèmes de calcul d'impact sont complémentaires et permettent d'évaluer plus finement les travaux des chercheurs.

5.5 Le peer-review

Ce processus permet, grâce à une relecture par des experts, anonymes ou non, soigneusement sélectionnés pour le domaine de recherche concerné, de donner une fiabilité et une crédibilité à des documents et des articles avant leur publication.

5.6 Le droit d'auteur

Les principes de base du droit d'auteur permettent de protéger les auteurs de l'usage non-autorisé de leurs œuvres, de définir les conditions de cette protection et de fixer les exceptions par rapport à cet usage. Ce droit est aussi un encouragement à l'investissement intellectuel. Le droit à la propriété intellectuelle fait partie du droit privé et englobe le droit d'auteur. En Suisse, la LDA (RS 231.1 Loi fédérale sur le droit d'auteur et les droits voisins), qui a été adoptée après plus de trente ans de travaux en 1992, définit qu'une œuvre est « une création de l'esprit qui a un caractère individuel et qui est de nature littéraire, scientifique ou autre ». Une révision est entrée en vigueur au 1er juillet 2008.

De manière générale, l'auteur d'une œuvre scientifique est propriétaire du copyright de la version "pré-publication" de son œuvre et il est libre de la diffuser sans autorisation. Par contre, une fois la publication validée, il ne peut plus décider librement de la diffusion de la version publiée. Il faudra qu'il négocie un accord de cession de copyright avec l'éditeur.

On trouve des informations complémentaires en anglais sur le droit d'auteur et le copyright dans l'aide du site d'Infoscience (http://wiki.epfl.ch/infoscience/copyright-faq-en).

[1] Informations extraites du site : http://hlwiki.slais.ubc.ca/index.php/Impact_factor#E-index

5.7 APIs

Il s'agit d'une interface de programmation fournie par un programme informatique. Elle permet à des programmes d'interagir les uns avec les autres. Aussi, ensemble de fonctions permettant d'accéder aux services d'une application, par l'intermédiaire d'un langage de programmation.

5.8 Les « Web services »

Ce sont des programmes informatiques qui permettent la communication et l'échange de données entre des applications et systèmes hétérogènes. Ces programmes sont accessibles sur le réseau internet et sont indépendants des systèmes d'exploitation et des langages de programmation. Ils privilégient l'interopérabilité et la simplification des échanges de données.

Ils permettent de traiter les données d'un poste-client en les important vers un serveur sur lequel tourne une application. Les ressources sont mises à disposition sur le web, d'où le nom "web service". Ce service fonctionne en définissant un format pour les requêtes et les réponses, ensuite le client envoie sa requête, le service web la traite, puis renvoie la réponse au client. Il existe un annuaire répertoriant les web services, disponible sous http://www.uddi.org.[2]

5.9 DOIs

C'est un identifiant associé à une ressource numérique. Il est le pendant numérique du numéro ISBN. Il sert à la citation d'objets sur internet. Sa pérennité est en général plus assurée que celle d'une adresse web (url).

5.10 ISSN

Il s'agit d'un numéro international normalisé permettant d'identifier de manière univoque une publication en série.

5.11 Journal scientifique (revue scientifique ou périodique scientifique)

C'est une publication académique et périodique destinée à signaler l'avancement des travaux de recherche dans un domaine particulier.

5.12 Métadonnée

C'est une donnée décrivant une autre donnée. Elle permet l'indexation des données, facilite la recherche, le partage et l'échange d'informations.

5.13 Norme ISO pour ISSN

Il s'agit d'unsemble de prescriptions techniques permettant de garantir le bon usage des identifiants de publications en série.

[2] Informations extraites de : http://www.gchagnon.fr/cours/xml/servicesweb.html

6 Consolider et compacter la base des journaux

6.1 Les outils

Une liste d'outils à évaluer était indiquée dans le cahier des charges. Il a fallu les analyser, puis évaluer leurs points forts et faibles et voir si des fonctionnalités seraient disponibles pour compléter les informations déjà présentes dans la base des journaux. J'ai établi un tableau comparatif avec l'url de référence de l'outil, une description succincte, les plus et les moins, des remarques, ainsi que l'éventuelle présence d'informations concernant:

- Le facteur d'impact
- Le droit d'auteur
- Les noms abrégés de journaux
- Le peer-review
- Web services
- APIs
- Prix (gratuit ou payant)

On peut classer les outils recensés en trois grandes catégories : les gratuits, les commerciaux et les académiques.

Un échantillon extrait de la base de journaux d'Infoscience a été sélectionné pour vérifier si on pouvait les retrouver dans les différents outils. Les résultats sont disponibles dans le second tableau ci-dessous.

Outil et url	Description	Facteur d'impact	Droit d'auteur	Noms abrégés	Peer-review	Web service	API	Prix	Couverture selon échantillon	Les plus	Les moins	Remarques
Ulrich's web ---------------- http://www.ulrichsweb.com/ulrichsweb/	Source d'autorité pour les informations bibliographiques et éditoriales Contient plus de 300,000 périodiques de tous les types (periodicals,academic and scholarly journals, Open Access publications, peer-reviewed titles, popular magazines, newspapers, newsletters, and more from around the world).	oui (lien sur JCR)	possible avec module payant	oui + variantes de titres	oui	oui (XML Data Service)	non	Payant. Abonnement EPFL	100%	- Très complet - Convivial et facile d'utilisation - Création d'alertes - Informations sur les mises à jour mensuelles de titres (ajout, modification, fusions,…) - Recherche simple, avancée ou par équations avec opérateurs booléens - Possibilités d'envoyer les résultats de recherche via email, de les télécharger ou de les imprimer	- Pas d'infos sur le droit d'auteur, mais possible en payant pour un module supplémentaire.	
Ulrich'sSAS ---------------- http://www.ulrichsweb.com/ulrichsweb/analysis/default.asp?navPage=4&	Permet d'identifier, d'analyser, d'évaluer et de créer des rapports concernant les publications en série sous forme papier et/ou électronique d'une institution.	oui (lien sur JCR)	non	oui (retour sur Ulrich's web)	oui	non	non	License acquise pour la bibliothèque. Usage interne et accès via compte spécial.		- Création de listes et outil de comparaison de bases de données. - Utile pour évaluer, mettre à jour et/compléter la liste de séries électroniques d'une bibliothèque.	- Vise plutôt les développeurs informatiques. - pas d'informations sur le droit d'auteur. - pas de web service, ni d'API	Payant, mais accès autorisé pour adminstrateurs de SISB
Scopus ---------------- http://www.scopus.com/home.url	Base de données de citations et de résumés issus de recherches bibliographiques et de sites Internet de qualité.	Journal metrics et RPM. Ainsi qu'un Journal Analyzer.	parfois	oui	oui (selon les titres)	non	oui (Scopus API)	Payant. Abonnement EPFL	90%	Fournit des listes d'informations (titres, éediteurs, Open Access) sous format excel. Mise à jour quotidienne. Impression de notices partielles ou intégrales, exportation d'informations dans un programme de gestion des références bibliographiques au format RIS ou ASCII, ajout à une liste personnelle et envoi par email.	- incomplet concernant le droit d'auteur. - pas de web service - peer-review pour 16'500 titres sur les 18'000 à disposition (http://www.info.sciverse.com/scopus/scopus-in-detail/facts/)	
ISI Web of Knowledge ---------------- http://www.isiknowledge.com/	Plateforme de recherché donnant accès à des bases de données de citations au niveau mondial	Oui via JCR	non	non	oui	non	non	Payant. Abonnement EPFL	52%	- Accès à un onglet "Additional Resources" listant des outils d'analyse et des sites web (listes de chercheurs, de noms d'organismes, de papiers de recherche, de sujets,…) - En lien avec Endnote	- Ne couvre pas toutes les publications. - pas d'informations sur le droit d'auteur. - pas d'informations sur les noms de journaux. - pas de web service, ni d'API	
Sherpa/ Romeo ---------------- http://www.sherpa.ac.uk/romeo/	Base de données produite par le "Centre de recherche en communication" de l'Université de Nottingham, au Royaume-Uni. Cet institut donne des informations sur les règles de publications édictées par les éditeurs concernant la politique d'archivage de leurs articles de journaux et de conférence sur le web et dans les dépôts d'archives institutionnelles de type Open Access. Collaboration avec JISC.	non	oui	non	oui	oui	oui	Gratuit	74%	- Documentation très compléte. - Romeo couvre des publications peer-reviewed et en série de plus de 700 éditeurs. - API permettant d'accéder aux différentes données (titres, issns, noms de contributeurs et couleurs) http://www.sherpa.ac.uk/romeo/api.html - Chaque entrée de la base fournit un résumé sur la politique générale de l'éditeur, la version et le lieu de dépôt de l'article, ainsi que les autres conditions attachées à ce dépôt. - Infos détaillées (copyright) par journal pour Chaque éditeur.	- Ne couvre pas toutes les publications. - pas d'informations sur le facteur d'impact. - pas d'informations sur les noms de journaux.	Donne des infos sur le droit d'auteur et l'archivage (classé par couleurs).

Outil et url	Description	Facteur d'impact	Droit d'auteur	Noms abrégés	Peer-review	Web service	API	Prix	Couverture selon échantillon	Les plus	Les moins	Remarques
xISSN (Web service) ---------------- http://xissn.worldcat.org/xissnadmin/index.htm	Web service qui fournit des informations sur les numéros ISSN des publications en série et leurs métadonnées. Ce service est disponible depuis Worldcat, le plus grand réseau de contenus et de services bibliographiques.	non	non	non	oui	oui	oui	Gratuit	97%	- Accès gratuit avec et sans inscription (inscription permet une consultation plus large du nombre d'issns). - Historique des numéros ISSN, via un outil gratuit permettant la génération de graphiques, disponible à partir de : http://worldcat.org/xissn/titlehistory - Outil de statistiques (nombre d'ISSNs et de groupes, alternate form, peer-reviewed,...) voir http://xissn.worldcat.org/xissnadmin/doc/stat.htm - API et web service (usage limité à 100 requêtes par jour, mise à jour mensuelle) disponibles sous: http://www.oclc.org/developer/services/xissn - Souscription possible à envois de mises à jour par email	- Base ne contient pas tous les issns attribués au niveau mondial (1 Issn électronique de l'échantillon n'est pas présent dans leur base). - pas d'email de contact (il faut aller sur la page de Worldcat) - pas d'informations sur le facteur d'impact. - pas d'informations sur le droit d'auteur. - pas d'informations sur les noms de journaux.	Donne des infos sur le numéro ISSN et ses variantes, le titre et le type de support de la publication, l'éditeur, les métadonnées et si peer-review.
JCR ---------------- http://isiknowledge.com/JCR	Ressource permettant d'évaluer et de comparer les données concernant les citations plus de 11'000 publications académiques et techniques (journaux les plus cités et journaux avec le meilleur taux d'impact par domaine…)	oui	non	oui	non	non	non	Payant. Abonnement EPFL	55%	- Divers types de mesures de citations des publications (Journal impact Factor, 5-Year Journal IF, Journal Self Cites, Journal Immediacy Index,…), visibles à la suite sous mode graphique. - Fournit aussi le rang du journal dans sa catégorie.	- Ne concerne pas tous les journaux. - pas d'informations sur le droit d'auteur. - pas d'informations pour le peer-review. - pas de web service, ni d'API.	Payant, mais accès autorisé pour EPFL.
issn.org ---------------- http://www.issn.org/	Source d'autorité pour l'attribution des numéros issn de ressources en série.	non	non	oui	non	non	non	Service payant et assez cher. Tarifs sous: http://www.issn.org/files/issn/bdc/BDC_en_2010.pdf	100%	- Recense tous les numéros issn au niveau mondial. - Informations très complètes sur les titres et leur évolution. - Eléments bibliographiques basés sur formats MARC 21, UNIMARC et ISSNMARC - Table ISSN_L qui regroupe les différentes versions d'une même ressource. - Documentation très complète	- Le prix de ce produit commercial. - pas d'informations sur le facteur d'impact. - pas d'informations sur le droit d'auteur. - pas d'informations sur le peer-review. - pas de web service, ni d'API	Service utilisé par MINERVA, SUDOC et EBSCO Réponse immédiate aux emails. Fourniture de comptes d'essai pour télécharger la table ISSN-L et rechercher 20 numéros issn.
SJR (SCImago) ---------------- http://www.scimagojr.com/	The SCImago Journal & Country Rank is a portal that includes the journals and country scientific indicators developed from the information contained in the Scopus® database (Elsevier B.V.). These indicators can be used to assess and analyze scientific domains.	oui avec SJR	non	non	oui	non	non	gratuit	68%	SCImago Journal Rank (SJR) indicator Infos tirées de la base Scopus In addition to available data at SCImago Journal & Country Rank, you can also embed significative journal metrics into your web as a clickable image widget.	- Ne couvre pas toutes les publications. - pas d'informations sur le droit d'auteur. - pas d'informations sur les noms de journaux. - pas de web service, ni d'API	Dans le champ de recherche, il faut saisir le numéro issn en omettant le tiret.

Outil et url	Description	Facteur d'impact	Droit d'auteur	Noms abrégés	Peer-review	Web service	API	Prix	Couverture selon échantillon	Les plus	Les moins	Remarques
BIMPE ---------------- http://bimpe.free.fr/bimpe_accueil.php	La "Base d'Information Mutualiste sur les Périodiques Electroniques" permet de récolter des informations sur des bouquets de périodiques numériques et de récupérer rapidement l'ensemble, ou une partie, des données essentielles d'un bouquet. Elle permet également de suivre les mises à jour qui interviennent au sein de ces différents bouquets.	non	non	non	non	non	non	gratuit	61%	- Mise à jour régulière et récente (dernière date 27 août 2010) avec possibilité de s'abonner aux mises à jour. - Infos complètes sur les titres répertoriés (statut, sigle, issn et issn_e, type d'accès, notes de filiation…) - Base interrogeable via un formulaire web.	- Base incomplète, tant au niveau des ISSNs que des informations fournies. - pas d'informations sur le facteur d'impact. - pas d'informations sur le droit d'auteur. - pas d'informations sur les noms de journaux. - pas de peer-review.	Fournit le sigle du journal et non pas le nom abrégé.
Périodiques de l'EPFL ---------------- http://library.epfl.ch/periodicals/	Tous les périodiques (journaux, revues scientifiques) disponibles à l'EPFL, sous formes électronique ou imprimée	non	non	oui (acronymes)	non	non	non	gratuit	81%	- Légende explicative pour types d'accès aux périodiques (libre, EPFL autorisé, mot de passe, commande d'article) - Donne les règles d'accès/utilisation des périodiques sous abonnement à l'EPFL (fichiers pdf ou html)	- Ne concerne que les périodiques accessibles à l'EPFL - Peu d'informations complémentaires, rien sur le facteur d'impact, ni le peer-review. - pas de web service, ni d'API	

Echantillon d'analyse pour critères d'enrichissement de la base des journaux													
No ID	Nom du journal	ISSN (print)	ISSN (elect	xISSN	Sherpa/	Ulrich's	EPFL period.	BIMPE	JCR	ISSN.c	Web of Sience	Scopus	SCImago
1	IEEE Pervasive Computing	1536-1268		oui	oui	oui	oui	oui	oui	oui	oui (titr	oui	oui (issn sans tiret)
			1558-2590	oui	oui	oui	**non**	**non**	**non**	oui	**non**	**non**	**non**
8	Journal of Physical Chemistry C	1932-7447		oui	oui	oui	oui	oui	oui	oui	oui	oui	**non**
			1932-7455	oui	oui	oui	oui	oui	**non**	oui	oui	oui	oui (titre ou e-issn)
21	International Journal of Business Process Integration and Management	1741-8763		oui	**non**	oui	**non**	**non**	**non**	oui	**non**	oui	**non**
			1741-8771	**non**	**non**	oui	**non**	**non**	**non**	oui	**non**	**non**	oui (titre ou e-issn)
66	Bulletin- Academie Nationale de Medecine	0001-4079	NULL	oui	oui	oui	**non**	oui	oui	oui	**non**	oui	oui (issn sans tiret)
90	CHINESE JOURNAL OF GEOPHYSI	0001-5733	NULL	oui	**non (pa**	oui	**non**	**non**	oui	oui	oui (titr	oui	oui (issn ou titre=Act
141	Agricultural Meteorology	0002-1571	NULL	oui	**non (tr**	oui	oui (issr	oui (IS	oui (I	oui		oui (anc	**non**
	Nouveau titre depuis 1985: Agricultural and Forest Meteorology	*ISSSN actuel=0168-1923)*											
207	American Journal of Sociology	0002-9602	NULL	oui	oui	oui	oui	oui	**non**	oui	oui	oui	oui (titre)
364	Astronomische Nachrichten	0004-6337	1521-3994	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui (titre ou e-issn)
413	AT&T Bell Laboratories Technical Journal	0005-8580	NULL	oui	**non**	oui	oui	**non**	**non**	oui	oui	oui	**non**
421	Biochemical Genetics	0006-2928	1573-4927	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui (titre ou e-issn)
725	Documentaliste	0012-4508	NULL	oui	oui	oui	oui	oui	**non**	oui	**non**	oui	oui (titre)
766	Electric Technology		0013-4155	oui	**non**	oui	oui	**non**	**non**	oui	**non**	oui	**non**
1180	Journal of Applied Probability	0021-9002	NULL	oui	oui	oui	oui	oui	oui	oui	oui (titr	oui	oui (titre)
1373	Journal of Pharmaceutical Sciences	0022-3549		oui	oui	oui	oui	oui	oui	oui	oui	oui	oui (issn sans tiret)
			1520-6017	oui	oui	oui	oui	oui	**non**	oui	oui	oui	oui
1413	Journal of Small Animal Practice	0022-4510	NULL	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui (titre)

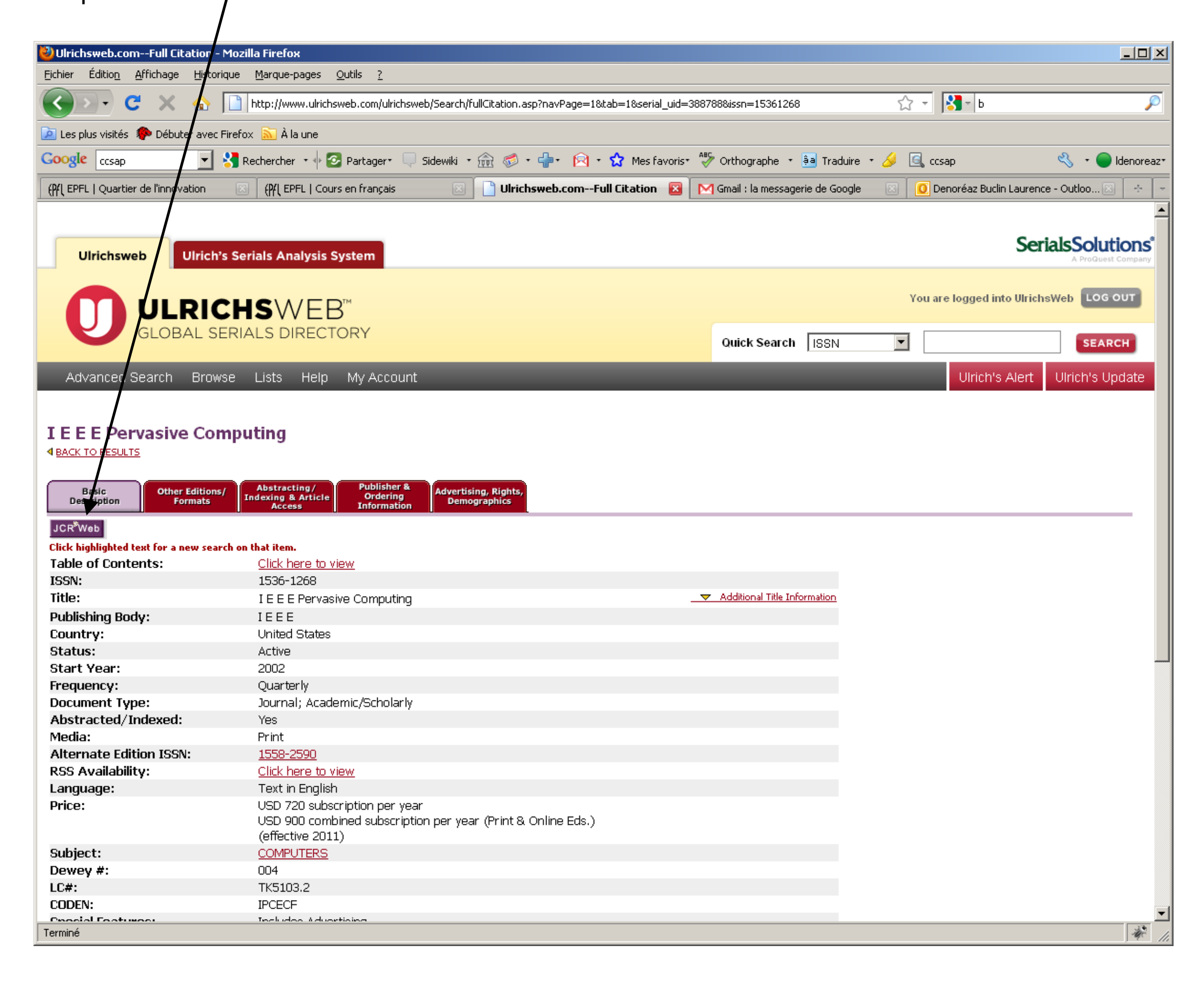
Echantillon d'analyse pour critères d'enrichissement de la base des journaux													
No ID	Nom du journal	ISSN (print)	ISSN (elect	xISSN	Sherpa/	Ulrich's	EPFL period.	BIMPE	JCR	ISSN.c	Web of Sience	Scopus	SCImago
1458	Keramische Zeitschrift	0023-0561	NULL	oui	oui	oui	oui	**non**	**non**	oui	**non**	oui	oui (issn sans tiret)
1686	Nuclear Technology	0029-5450	NULL	oui	oui	oui	oui	**non**	oui	? (oui)	oui	oui	oui (issn sans tiret)
1704	Proceedings- Nutrition Society of London	0029-6651	NULL	oui	oui	oui	oui	**non**	oui	? (oui)	**non**	oui	**non**
1710	Oesterreichische Ingenieur Zeitschrift	0029-9219	NULL	oui	**non**	oui	oui	**non**	non	? (oui)	**non**	oui	**non**
1847	Population Index	0032-4701	NULL	oui	oui	oui	oui	oui	non	? (oui)	**oui**	oui	**non**
3852	Philosophical Transactions of th	0261-0523	NULL	oui	oui (ave	oui	oui	oui	oui (is	? (oui)	**non**	oui (issı	oui (titre issn= 1364-5
3994	International Journal of Materia	0268-1900	NULL	oui	oui	oui	**non**	**non**	oui	? (oui)	**non**	oui	oui (issn sans tiret)
4069	Language and Communication	0271-5309	NULL	oui	oui	oui	oui	oui	**non**	? (oui)	**non**	oui	oui (issn sans tiret)
4090	Risk Analysis	0272-4332	1573-9147	oui	oui	oui (pas	oui	oui	oui (is	? (oui)	**non**	oui	oui (titre)
7112	Geometric and Functional Analy	1016-443X	1420-8970	oui	oui (pa:	oui	oui	oui	oui (is	? (oui)	oui	oui	oui (issn=14208970)
7114	Stratégies énergétiques (nom d	biosphère	1016-8397	Oui. (T	**non**	oui (Titr	oui	**non**	**non**	? (oui)	**non**	**non**	**non**
9999	Ocean Dynamics	1616-7341	1616-7228	Oui	oui (pa:	oui	oui	oui	oui (is	? (oui)	oui	oui	oui (issn=16167341)
	Taux de présence dans l'outil			97%	74%	100%	81%	61%	55%	100%	52%	90%	68%

6.2 Analyse des outils

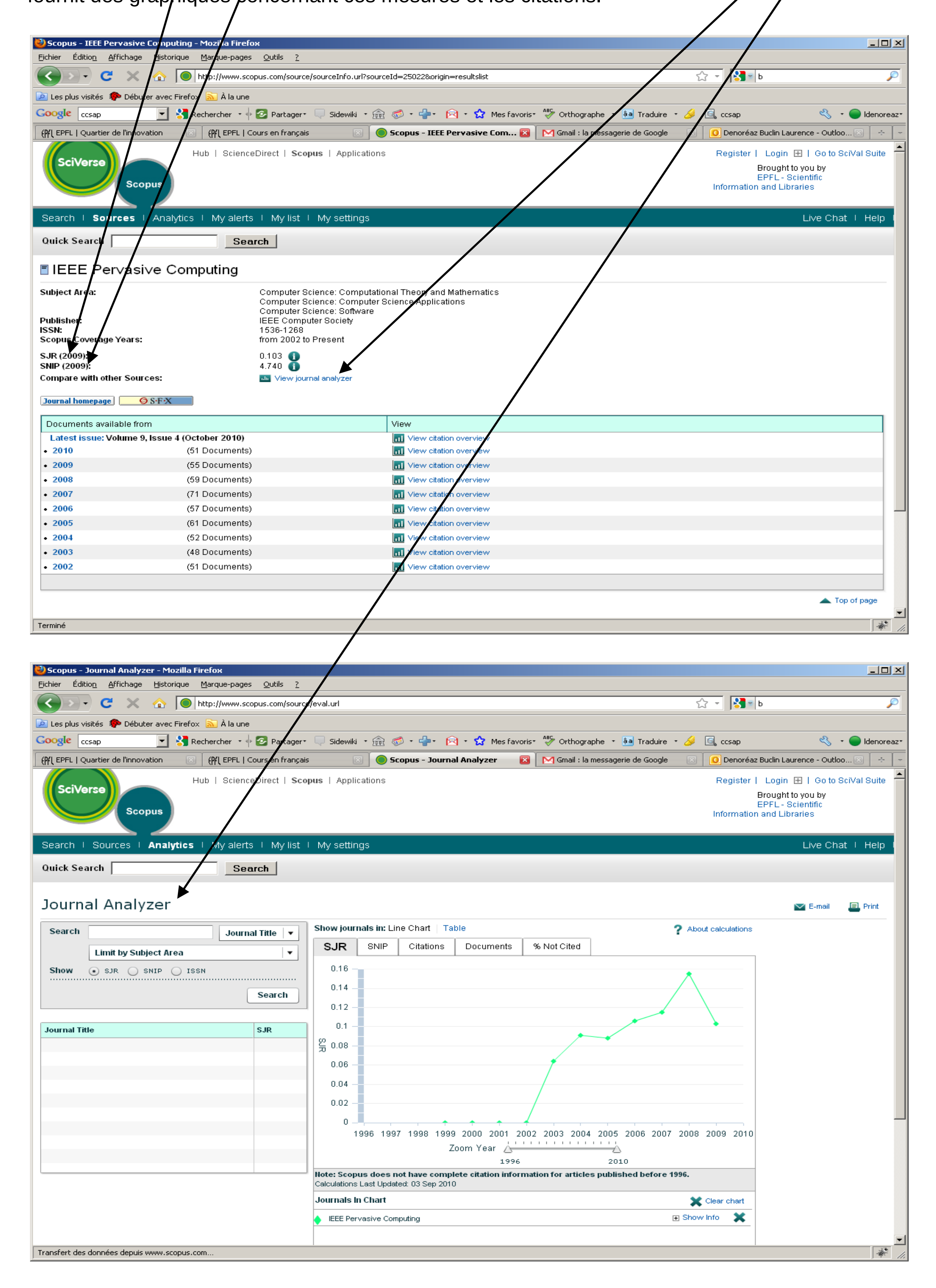
De cette analyse, il ressort que la Bibliothèque de l'EPFL a fait les bons choix tant au niveau des abonnements aux périodiques (fournisseurs d'informations fiables concernant les paramètres demandés), qu'au niveau de l'utilisation de Sherpa/Romeo pour la gestion des copyrights.

Pour la vérification des numéros ISSN et après la lecture d'un document sur SERVAL (http://archivesic.ccsd.cnrs.fr/sic_00540289/), j'ai contacté l'organisme qui les attribue : ISSN.org. Celui-ci a été d'accord de me donner deux accès temporaires, le premier pour télécharger leur table ISSN-L, « ISSN-L est un ISSN désigné par le réseau ISSN pour regrouper les différents supports physiques d'une même ressource continue » (http://www.issn.org/1-22637-Qu-est-ce-qu-un-ISSN-L-.php) et le deuxième pour accéder à leur portail et pouvoir vérifier dans leur base officielle mon échantillon de journaux. Par contre, la consultation a été limitée à vingt titres.

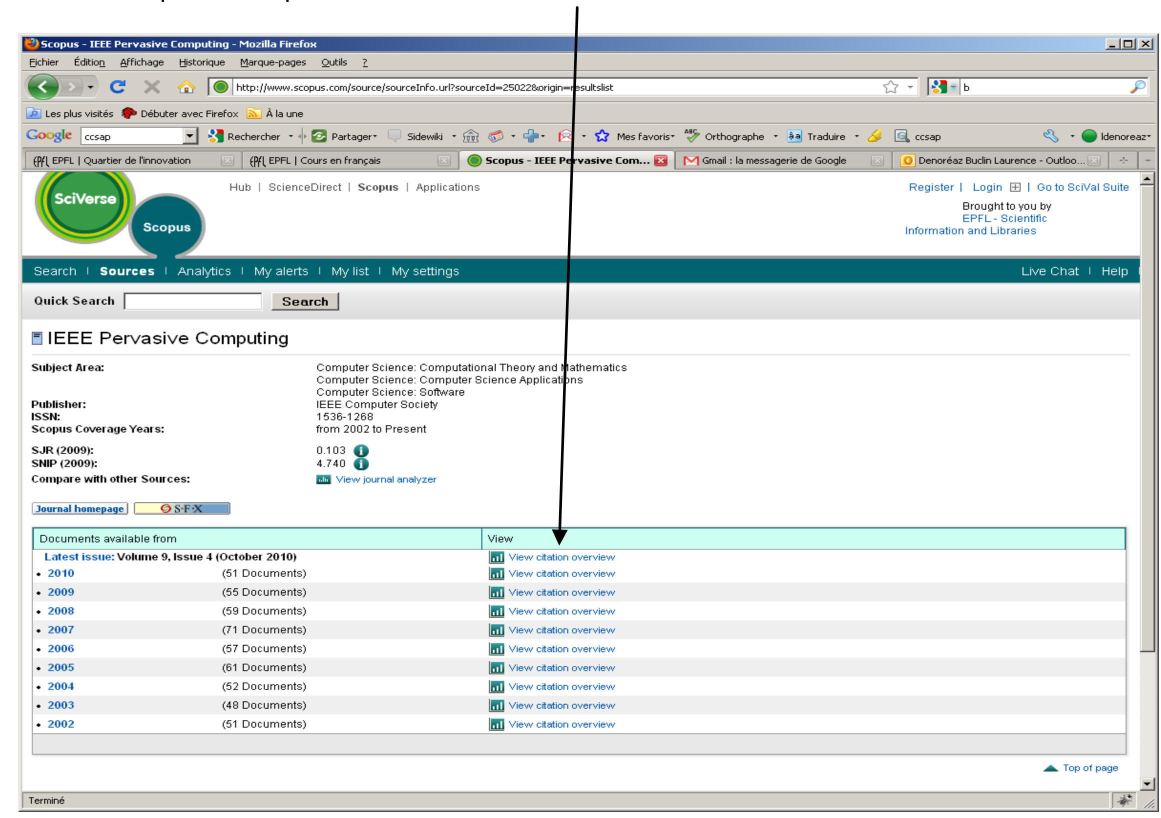
Ulrich's est vraiment l'outil le plus complet, tant en ce qui concerne les titres de périodiques, que pour les fonctionnalités supplémentaires disponibles, soit avec l'abonnement existant, soit avec un complément à celui-ci pour obtenir des informations sur le copyright. Il propose aussi un lien avec JCR pour les données sur le facteur d'impact, ce qui pourrait permettre de compléter les informations nécessaires dans Infoscience.



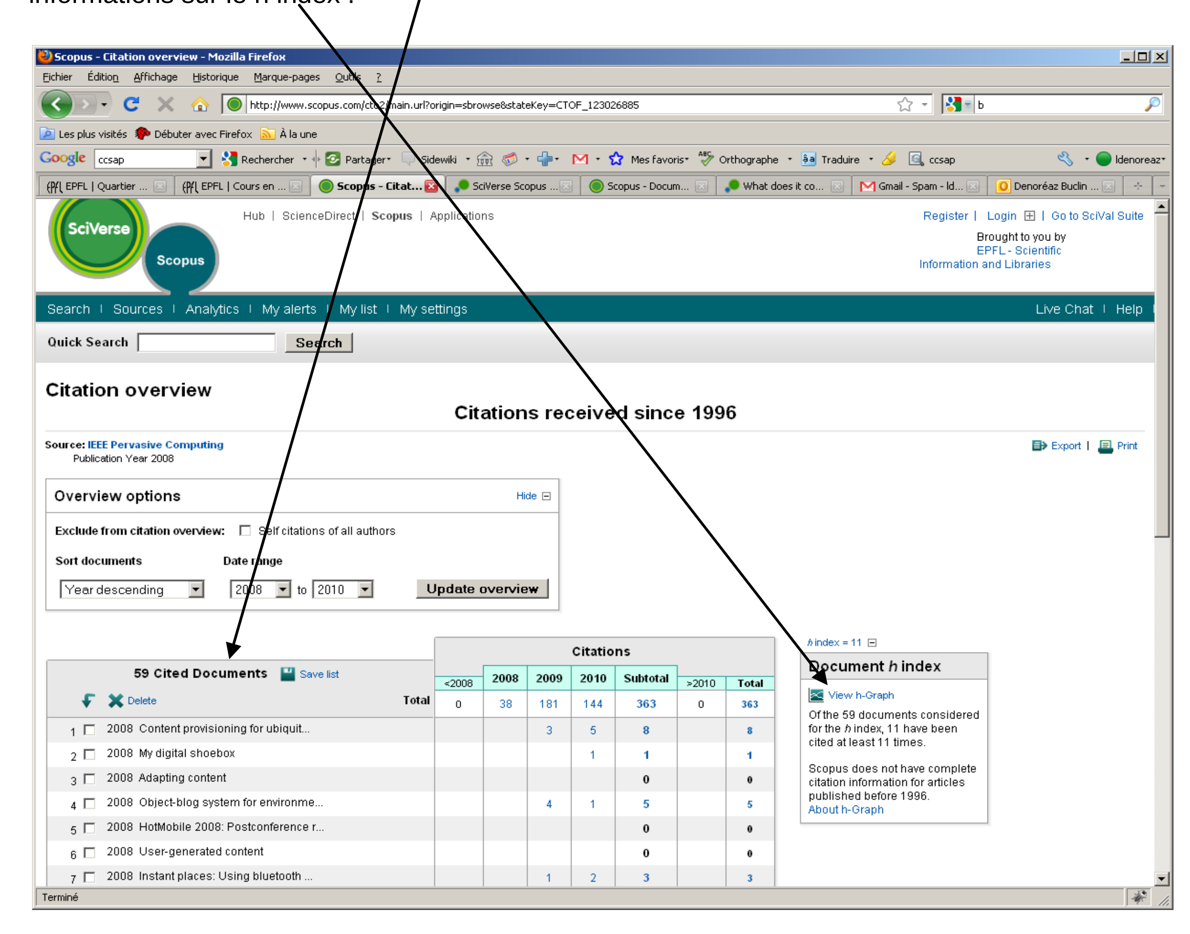
Scopus est bien équipé au niveau des fonctionnalités. Il renseigne en plus sur d'autres types de mesures (SJR et SNIP). Il permet aussi d'accéder à une fonction "Journal Analyzer" qui fournit des graphiques concernant ces mesures et les citations.



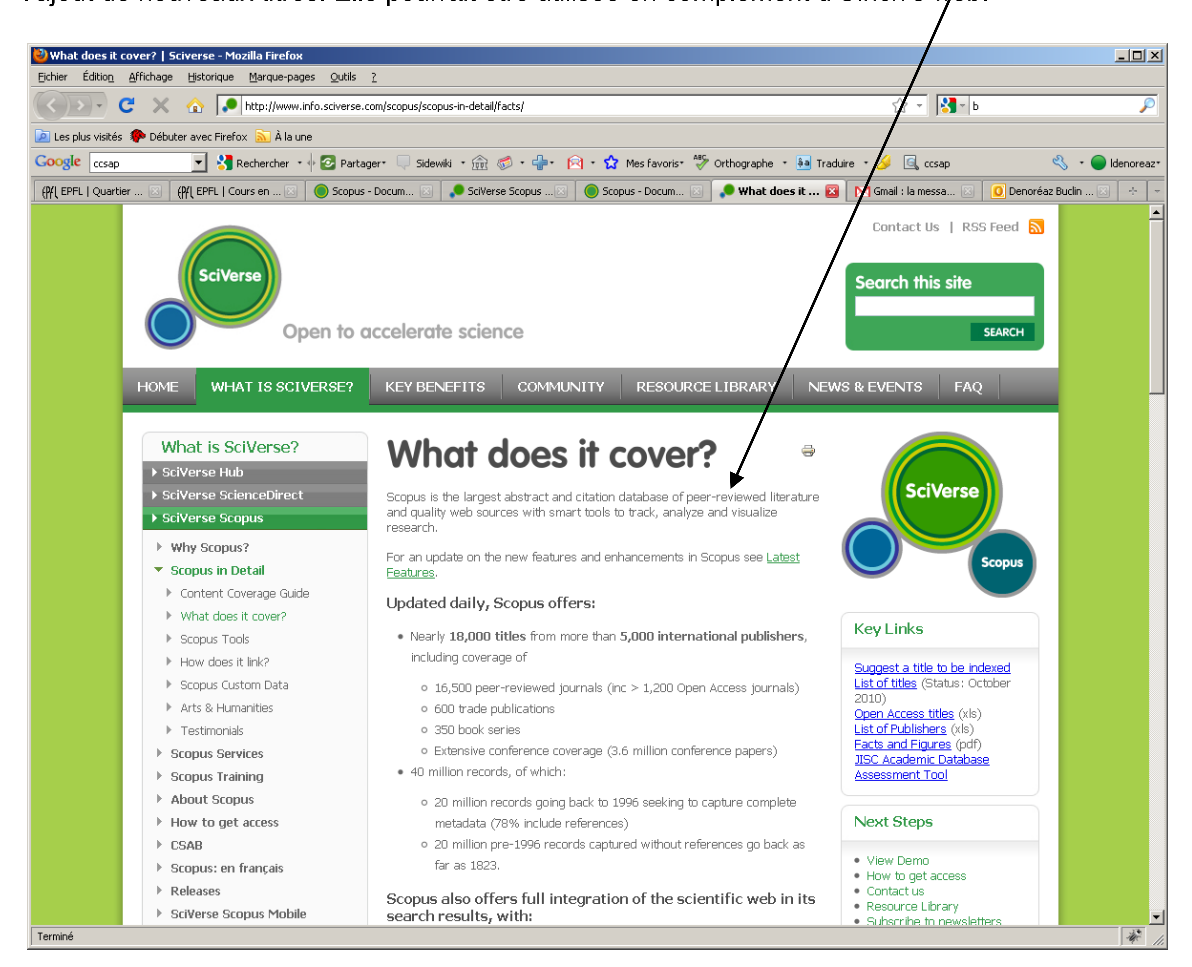
En allant cliquer sur la partie : View citation overview



on obtient le détail des citations pour le numéro du journal concerné, ainsi que des informations sur le h index :



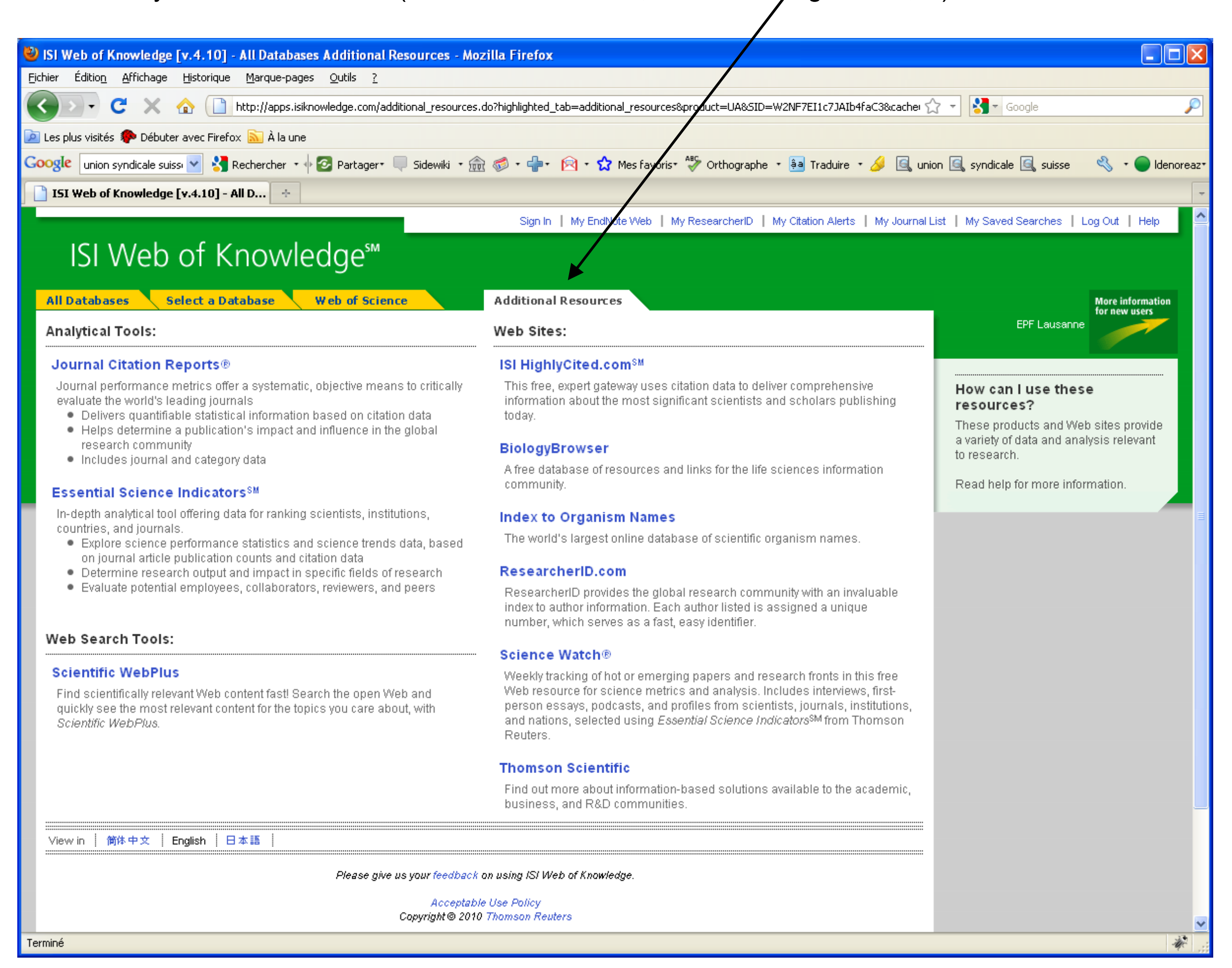
Scopus est également la plus grande base de données de littérature revue par les pairs et elle bénéficie d'une mise à jour quotidienne de ses données et d'une mise à jour annuelle pour l'ajout de nouveaux titres. Elle pourrait être utilisée en complément d'Ulrich's web.



Scopus permet de télécharger des informations complètes concernant les titres de publications (http://www.info.sciverse.com/sciencedirect/content/journals/titles).

Dans ISI Web of Knowledge, la documentation donne comme information que les journaux sont soigneusement sélectionnés, et en particulier que l'un des standards de sélection est le peer review (http://wokinfo.com/about/whatitis/ ou http://wokinfo.com/about/whatitis/) Par contre, le rythme de mise à jour des informations n'est pas mentionné.

ISI Web of Knowledge propose aussi un accès à des ressources additionnelles listant des outils d'analyse et des sites web (listes de chercheurs, de noms d'organisme,…) :



Les outils commerciaux mettent en général à jour leurs informations en permanence, donc il est possible de récupérer celles-ci chaque jour. La question à se poser est de savoir si on veut être à jour en temps réel ou s'il suffit de charger les données quelques fois dans l'année.

Malgré des recherches approfondies, je n'ai pas trouvé de web services ou API qui pourraient nous fournir facilement les renseignements demandés, autres que celles déjà connues ou utilisées par Infoscience.

La base des journaux d'Infoscience mériterait d'être mise à jour avec des informations récentes, en particulier pour les titres, mais aussi pour les numéros ISSN. En effet, dans l'échantillon de journaux, on remarque que certaines informations datent quelque peu.

Pour conclure, je dirai que les outils de volontariat sont moins valables, malgré le fait qu'ils offrent leurs services en général gratuitement et sans restriction. BIMPE fait un effort très intéressant, mais avec des moyens limités, donc les résultats sont aussi limités, et leurs données ne nous donnent pas tous les renseignements escomptés : il manque souvent des numéros ISSN ou d'autres renseignements.

A l'opposé, le JISC financé par le gouvernement anglais a des ambitions d'excellence au niveau mondial concernant l'enseignement et la recherche. Cet organisme est très actif en Europe pour la communication et la publication académiques, ainsi que pour la structuration et et les stratégies touchant aux publications. Les outils académiques peuvent être un bon compromis, mais cela demande beaucoup de temps et d'énergie pour joindre la bonne personne et créer les bons contacts avec l'institution.

En définitive, ce sont bien les outils payants qui fournissent les informations les plus exhaustives et les plus fiables.

6.3 Stratégies pour l'ajout d'informations

Les stratégies proposées sont les suivantes :

Pour le rythme de mise à jour général des informations : 2 fois par année serait le minimum conseillé. En fait, cette fréquence dépend grandement du temps de chargement des informations dans Infoscience.

Pour chacun des paramètres, il est pertinent d'aller récupérer les informations sur l'outil le plus fiable :

- pour le facteur d'impact, on utilisera JCR de ISI Web of Knowledge, éventuellement à partir d'Ulrich's web. Scopus et éventuellement SCImago permettront de compléter les données concernant la bibliométrie.

- pour les numéros ISSN, le peer-review et les noms abrégés de journaux, l'outil fournissant les meilleurs renseignements est à nouveau Ulrich's web, Scopus le suivant de près par rapport à l'exhaustivité des renseignements.

- pour les ISSN spécifiquement, l'utilisation de xISSN, le web service de Worldcat, est tout à fait indiquée. Il faudrait éventuellement étudier plus finement, la possibilité de souscrire un abonnement chez ISSN.org, l'organisme qui attribue les numéros ISSN pour le monde entier.

- pour le droit d'auteur, Sherpa/Romeo est bien l'outil qui donne les meilleures indications. En cas de doute, il est conseillé de se renseigner sur le site de l'éditeur ou en prenant contact avec lui. D'autres dépôts utilisent par exemple les services payants de Rightsphere, un produit de la firme Copyright (http://www.copyright.com/viewPage.do?pageCode=bu11), elle-même partenaire de Ulrich's.

- pour les noms abrégés de journaux, il faudra à nouveau se baser sur les données fournies par Ulrich's, Sopus et JCR.

6.4 Les autres dépôts d'archives institutionnelles

Dans le but de voir comment cela se passe dans les autres dépôts d'archives institutionnelles et d'éventuellement partager des informations, j'ai eu un contact par email avec certains d'entre eux. Pour les choisir et avoir un éventail large du point de vue géographique, j'ai consulté les institutions listées dans le classement Webometrics (http://repositories.webometrics.info/), puis j'ai fait ma sélection en Suisse, en Europe et en Amérique du Nord.

Une partie d'entre eux ne m'a pas répondu, quelques-uns n'ont pas voulu entrer en matière par manque de temps et un très petit nombre a répondu favorablement à ma demande en envoyant des réponses à mes questions.

La réponse la plus complète et la plus cordiale m'est parvenue du dépôt Alexandria de l'Université de St-Gall (http://www.alexandria.unisg.ch/). Il s'agit d'une plate-forme sur laquelle les chercheurs entrent eux-mêmes les informations, par ex pour le paramètre du "peer-review", ils font confiance aux données entrées par le chercheur. Idem pour le copyright, seuls les full-texts sont soumis à une validation par "disclaimer ". Concernant les web services, ils font appel à l'API d'Amazon pour les couvertures de livres. Et leurs numéros ISSN sont extraits à partir de plusieurs sources (VHB-Jourqual, journal websites et Ulrich's web). Ils ont le projet d'améliorer toutes leurs données et d'ajouter de nombreuses informations complémentaires pour la fin de l'année 2010.

Les autres dépôts ayant joué le jeu et m'ayant transmis quelques informations sont Archimer (http://archimer.ifremer.fr/)et OATAO (http://oatao.univ-toulouse.fr/).

Parmi les dépôts les plus complets et les plus actifs, on peut citer **Hal** (http://hal.archives-ouvertes.fr/), le site des archives ouvertes du CNRS et **ORBI** de l'université de Liège (http://orbi.ulg.ac.be/)**.** Même si le premier n'a pas répondu à ma demande d'informations et que le second n'a pas été contacté, car M. Aymonin, le directeur de la Bibliothèque de l'EPFL est déjà en contact avec eux.

Il ressort de cette analyse que les institutions ayant répondu sont celles qui connaissent Infoscience et qui ont déjà eu un échange de contacts, soit par le biais par exemple d'une conférence, soit en connaissant déjà l'EPFL. .Les autres universités ne semblent pas intéressées par une grande Ecole dans un si petit pays, tet que la Suisse.

Le bilan pour cette partie du travail est plutôt mitigé. Il semblerait qu'à l'heure du libre accès, une grande partie des institutions ne soient pas vraiment prêtes à collaborer.

6.5 Utilisation des informations trouvées sur les sites des outils

Comme précisé sur les différents sites (par exemple Ulrich's : http://www.ulrichsweb.com/ulrichsweb/Registration/termsOfUse.asp), pas d'utilisation, ni de reproduction des informations sans autorisation préalable pour tous les outils payants, sauf pour un usage non-commercial, scolaire ou à but de recherche. Idem, en général pour les outils académiques. Seuls les outils de type mutualiste ou volontaire permettent d'utiliser librement leurs données.

6.6 Lectures complémentaires

Quelques lectures complémentaires m'ont permis de mieux appréhender l'importance de la bibliométrie et des articles scientifiques. Elles sont listées dans la bibliographie à la fin du rapport.

En particulier, ces lectures m'ont conforté dans l'idée que le facteur d'impact n'est pas une mesure suffisante pour l'évaluation de la littérature scientifique.

Elles m'ont aussi permis de découvrir d'autres portails moins connus référençant les périodiques et les articles.

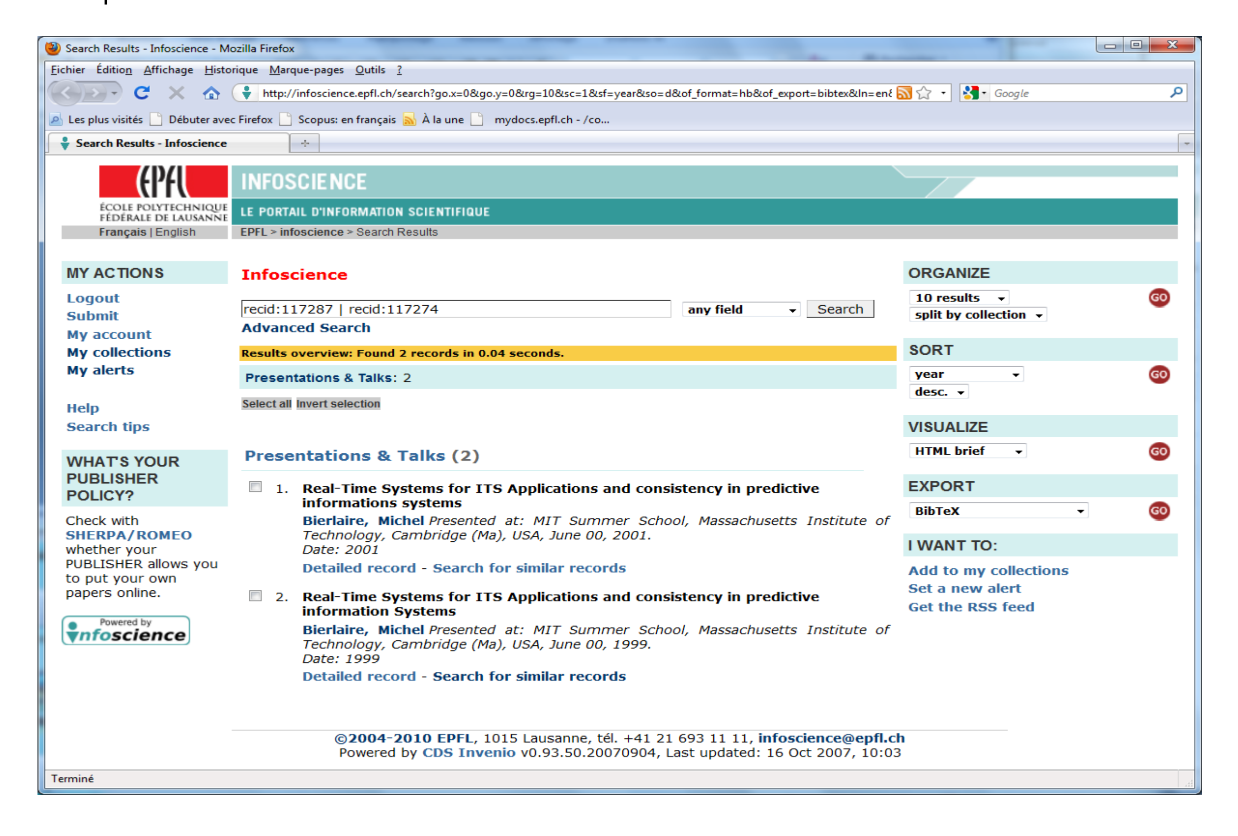
Surtout, elles m'ont montré l'ampleur des collaborations et des intérêts financiers en jeu entre les différents fournisseurs d'information sur le web. On retrouve parmi ces prestataires Google, Thomson Reuters, Elsevier,...

7 Interface de déduplication

Pour cette partie du travail, il s'agit de trouver des solutions pour traiter les documents saisis dans le dépôt d'archives institutionnelles de l'EPFL et se ressemblant fortement (doublons, triplets,…) en proposant une méthode d'évaluation et de traitement de ces documents, ainsi qu'une interface de déduplication. Cette interface devrait aussi permettre de traiter les éventuels doublons provenant de bases de données externes (Web of Science).

Mi-juillet 2010, il y a environ 2'000 documents existants à double dans l'archive institutionnelle de l'EPFL et qui doivent être évalués individuellement par les bibliothécaires scientifiques. Un doublon peut être un vrai doublon, pour lequel tout est identique, sauf par exemple la liste des auteurs ou leurs noms (orthographe différente, majuscules/minuscules) ou éventuellement d'autres informations que l'on pourra compléter en les transférant d'une notice à l'autre. Il peut aussi s'agir d'un faux doublon, donc de deux objets différents qu'il faudra conserver séparément.

Un exemple de doublon :



7.1 Références pour Infoscience

- Le site web pour consultation et soumission de données (http://infoscience.epfl.ch)
- Le cahier des charges de la plateforme de gestion bibliographique de l'EPFL, nommée « Curator » datant de février 2009 et prévoyant les développements futurs de ce dépôt (http://infoscience-wiki.epfl.ch/Curator?action=AttachFile&do=get&target=curator-CdC-v1.6.pdf)
- Le wiki de travail et d'informations (http://infoscience-wiki.epfl.ch/)

7.2 Gestion des doublons

La partie 6 du cahier des charges (Curator) explique les différentes étapes prévues pour le traitement, l'évaluation et l'éventuelle suppression de notices existant en plusieurs

exemplaires. Elle précise qui pourra exécuter des mises à jour des documents saisis, ainsi que la manière de gérer certains cas particuliers (notices à deux propriétaires et faux positifs).

7.3 Méthode de travail

Pour aborder la problématique du traitement des doublons dans Infoscience et mieux cerner le fonctionnement de ce dépôt, je me suis familiarisée avec la consultation des données internes et la saisie de nouvelles entrées. Cela m'a donné une vue précise des différentes fonctionnalités de l'interface d'Infoscience, de ses possibilités de tri et d'exportation des notices extraites, des différents formats (HTML brief/detailed/Marc) et des différents champs d'une notice (type de doc, titre, auteur, labo, abstract, full text,…).
J'ai obtenu le statut « administrateur » me permettant de gérer les données de la base officielle, ainsi qu'un accès privilégié à la base de test.
J'ai également visualisé un certain nombre de doublons provenant du fichier mis à ma disposition dans l'optique de repérer et comprendre les différents cas de figure.

7.4 Rencontres avec les bibliothécaires scientifiques

Pour mieux appréhender les problèmes spécifiques des bibliothécaires, j'ai eu un entretien avec trois d'entre eux. Ceci m'a permis d'identifier leurs besoins et de pouvoir élaborer un processus, en vue de proposer une interface de gestion des doublons.

Chacun d'eux m'a donné un certain nombre d'informations sur les différents cas de doublons, leur manière de les évaluer et de gérer ces entrées multiples, ainsi que leurs souhaits pour faciliter cette gestion. J'ai établi deux listes (méthode et suggestions), visibles dans le tableau ci-dessous.

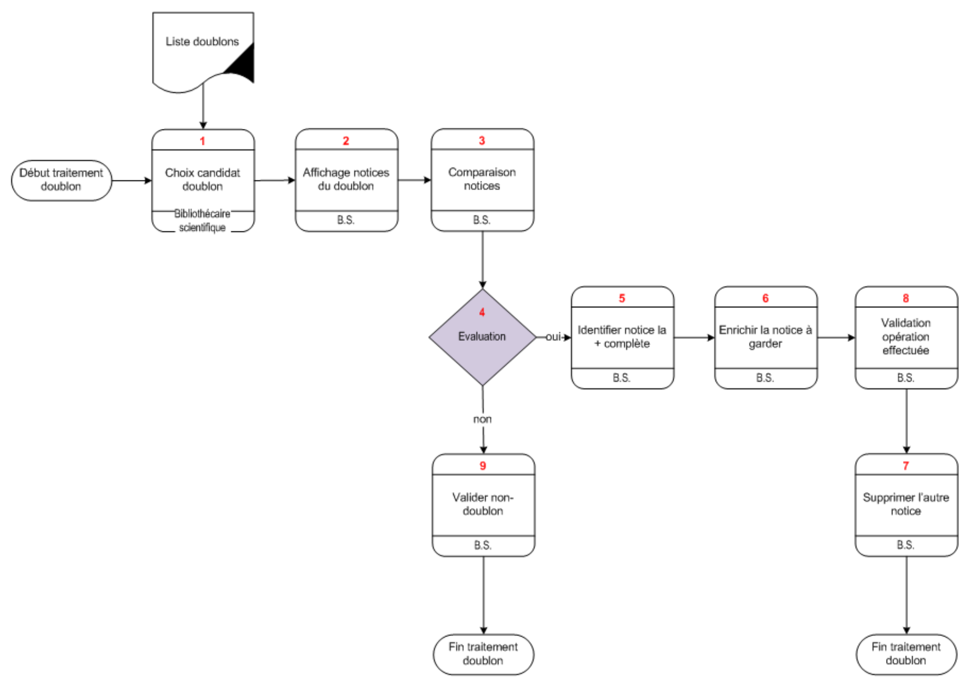
Méthode et évaluation	Suggestions d'amélioration
- Transmission du fichier contenant la liste des doublons, deux par deux, répartis par facultés, avec doublons classés par laboratoire. - Copier/coller dans l'interface Infoscience, de chaque ligne du fichier. - Affichage du html détaillé de chaque notice dans une fenêtre distincte. - Comparaison des différents champs (titre, auteur, année, no de pages et nom du laboratoire…). - Evaluation, s'il s'agit d'un vrai ou faux doublon. - Mise à jour dans BibEdit (suppression, transfert d'informations, conservation des deux,…) - Eventuellement, complément des notices existantes avec les infos d'autres bases de données.	- Affichage direct en HTML detailed - Mise en couleurs des champs communs/différents - Bouton de suppression direct pour chaque notice détaillée - Dans BibEdit, bouton pour ajout d'un nom de laboratoire - Possibilité de pouvoir basculer les champs plus complets dans les notices conservées - Possibilité d'ajouter des infos manquantes (par ex depuis PubMed, WOS,…) - Système de vérification des doublons déjà supprimés (mémoire)

Ils m'ont aussi fourni des explications supplémentaires concernant la gestion spécifique de certains types de documents (articles de revue, rapports d'activité, articles de conférence, posters,…).

Un souci exprimé par chaque bibliothécaire est le fait que BibEdit garde dans son cache les informations de la notice précédente, ce qui crée des problèmes pour la mise à jour des informations.

Après ces entretiens, j'ai fait le bilan de leurs demandes et besoins. Puis en collaboration avec Grégory Favre j'ai évalué ce qu'il était possible de mettre en place pour eux.

Pour clarifier la situation, j'ai préparé un schéma (fichier Visio ci-dessous) avec les différentes étapes du processus idéal de gestion des doublons.



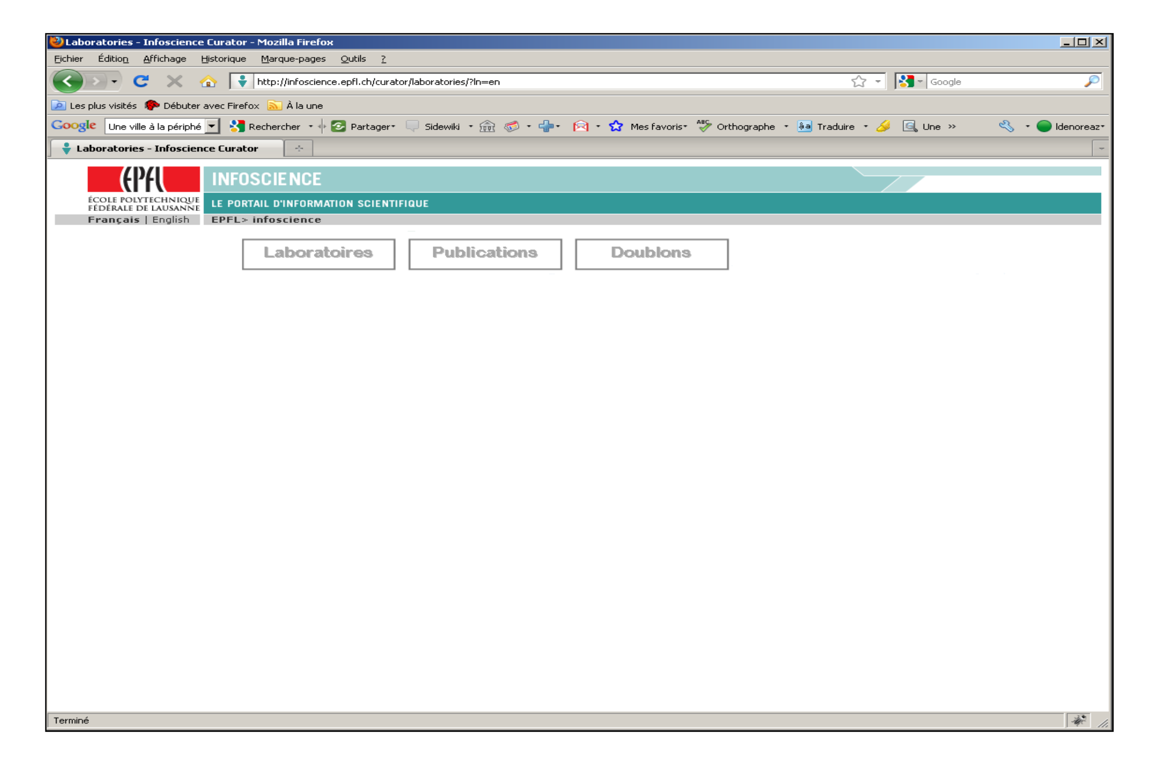
Description du processus :

Après avoir accédé à la liste des doublons, qui peut être affichée soit par faculté, soit par laboratoire ou complète (attention aux entrées mixtes sur deux facultés), le bibliothécaire scientifique (B.S.) choisit un candidat doublon (1). Les notices correspondantes s'affichent en parallèle (2). Il peut maintenant comparer les notices en mode d'affichage "HTML detailed" (3) et évaluer si il s'agit d'un vrai ou d'un faux doublon (4). Si c'est un vrai doublon, il faut identifier la notice la plus complète (5), si nécessaire en passant par le mode BibEdit. Il est aussi possible de procéder à l'enrichissement d'une des notices (notice forte) avec des informations de l'autre (6). Une fois l'enrichissement terminé, il faut valider les opérations effectuées (7). Ensuite, il convient de repasser en mode "HTML detailed" pour supprimer la notice déterminée comme faible (8). S'il ne s'agit pas d'un doublon (9), on validera chacune des notices en la déclarant à conserver.

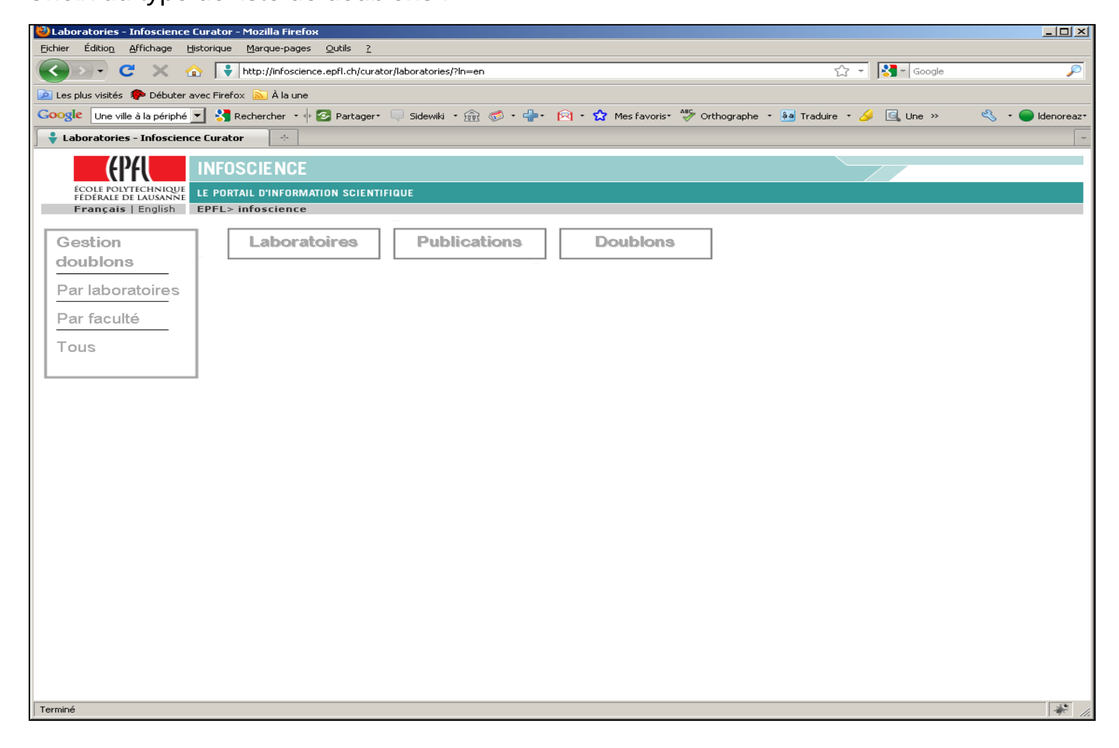
7.5 Proposition d'interfaces pour corriger les doublons

Ci-dessous se trouvent les écrans que je propose pour l'interface de déduplication réalisée en me basant sur les fonctionnalités demandées par les bibliothécaires scientifiques. La numérotation des écrans ne suit pas exactement celle des étapes du processus.

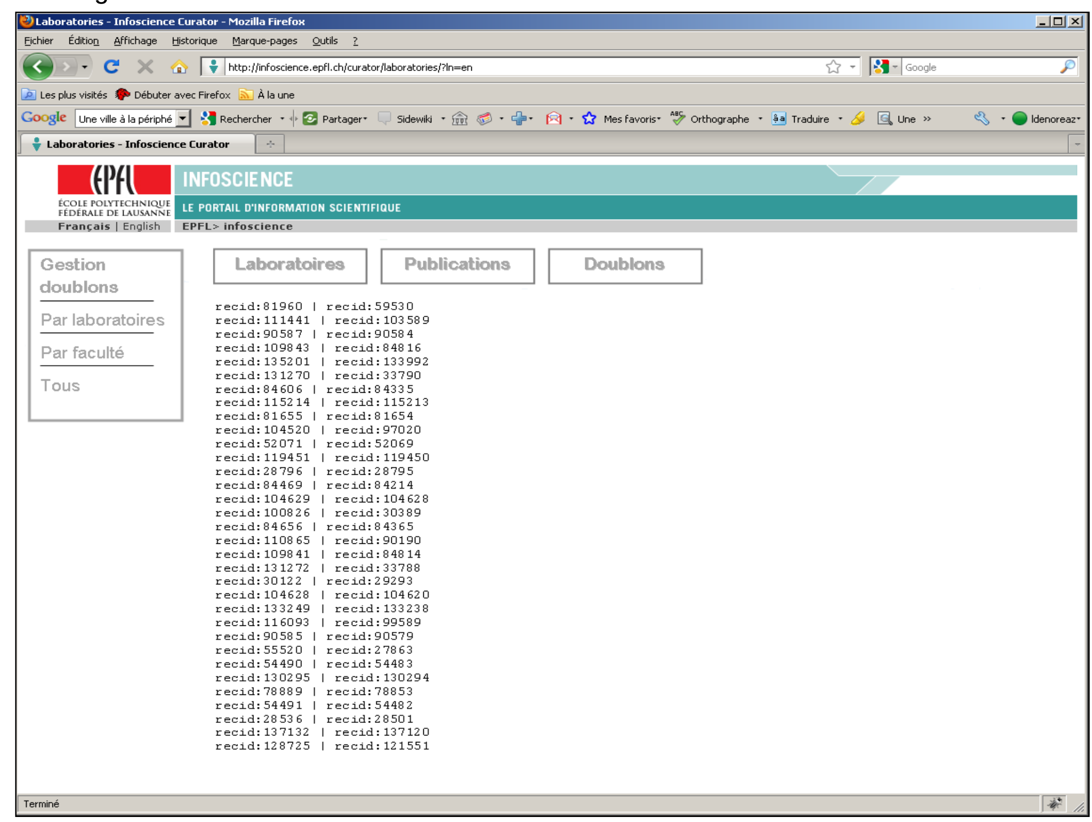
1. Accès à l'interface (disponible depuis http://infoscience.epfl.ch/curator/):



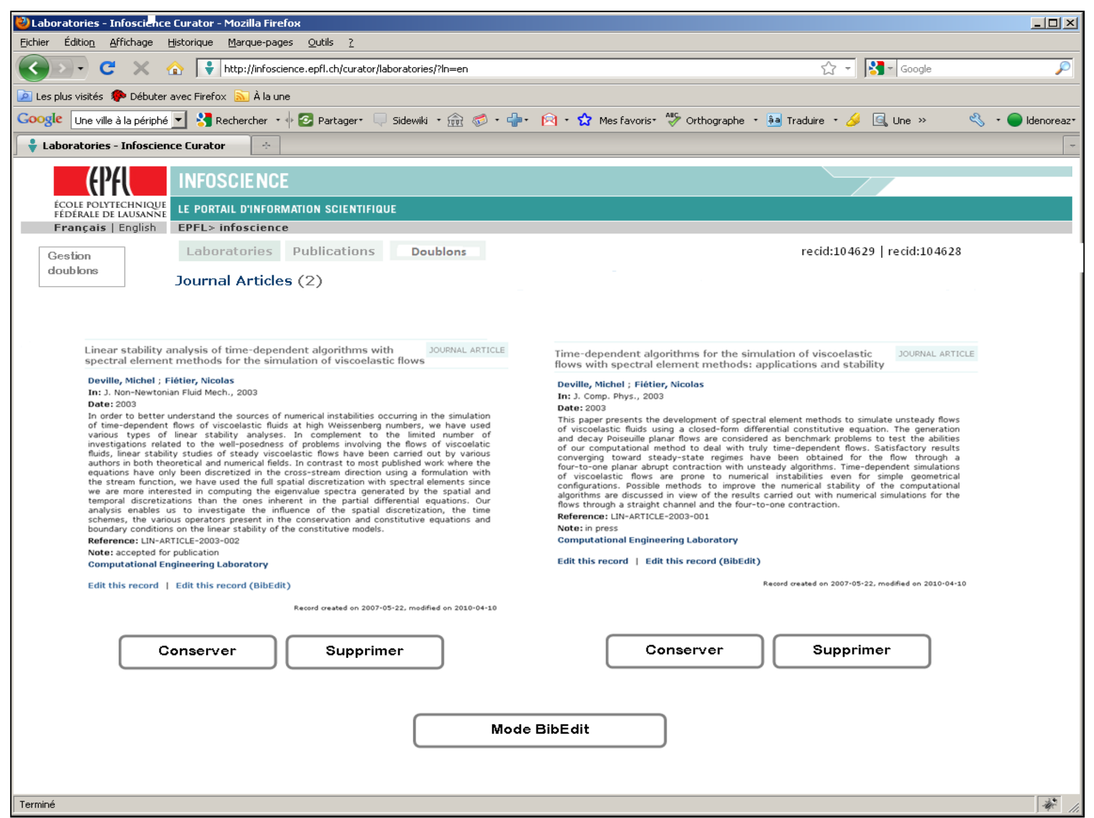
2. Choix du type de liste de doublons :



3. Affichage de la liste selon choix effectué:

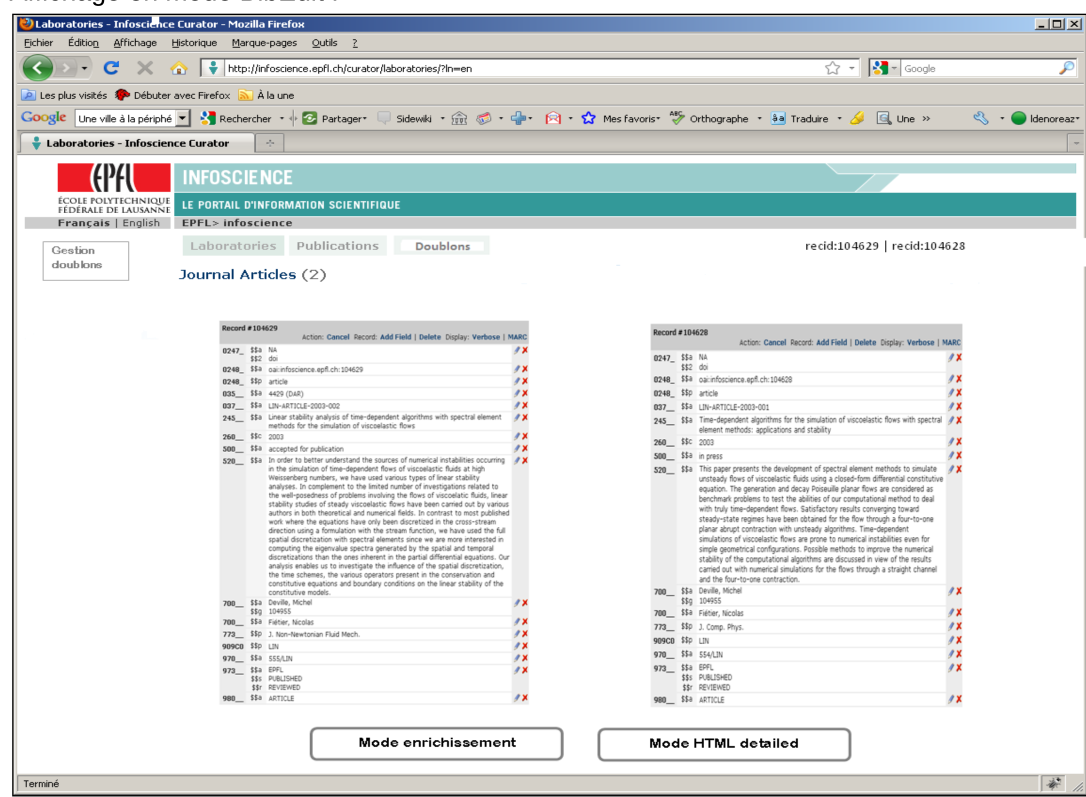


4. Sélection du doublon en cliquant sur la ligne correspondante et affichage en mode HTML detailed des 2 notices

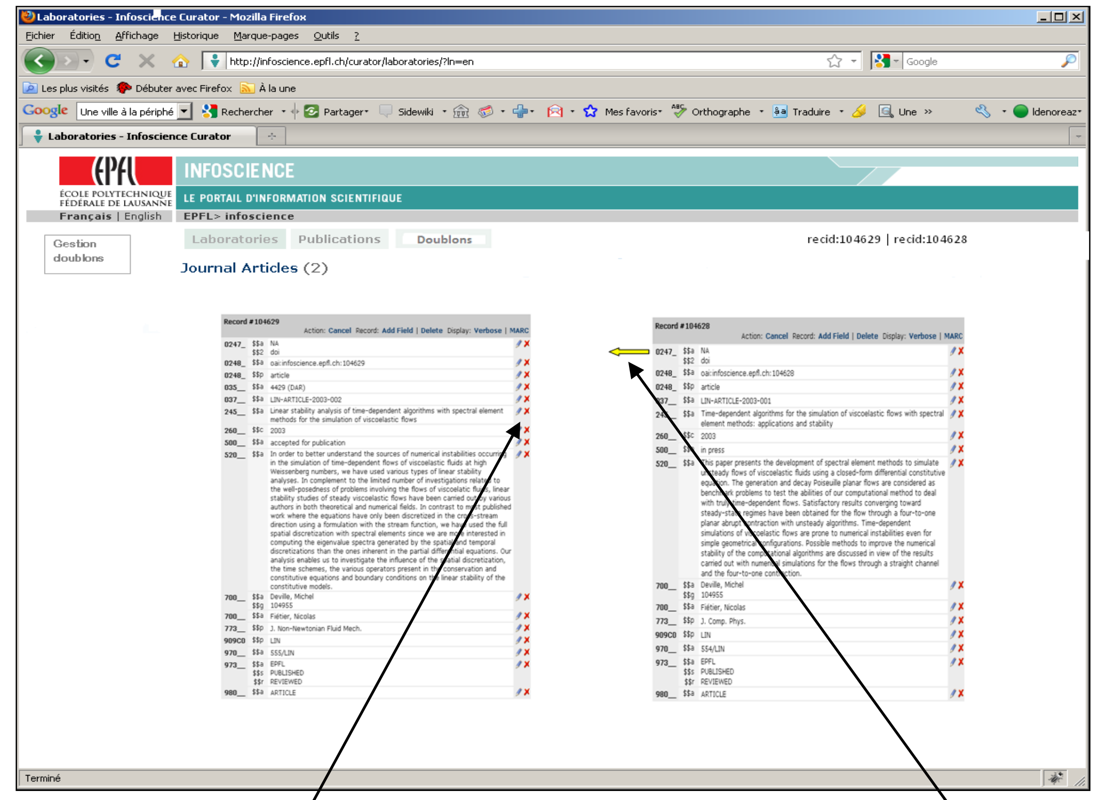


5. Evaluation/comparaison avec possibilité de conserver/supprimer une des notices ou de passer en mode BibEdit

6. Affichage en mode BibEdit :

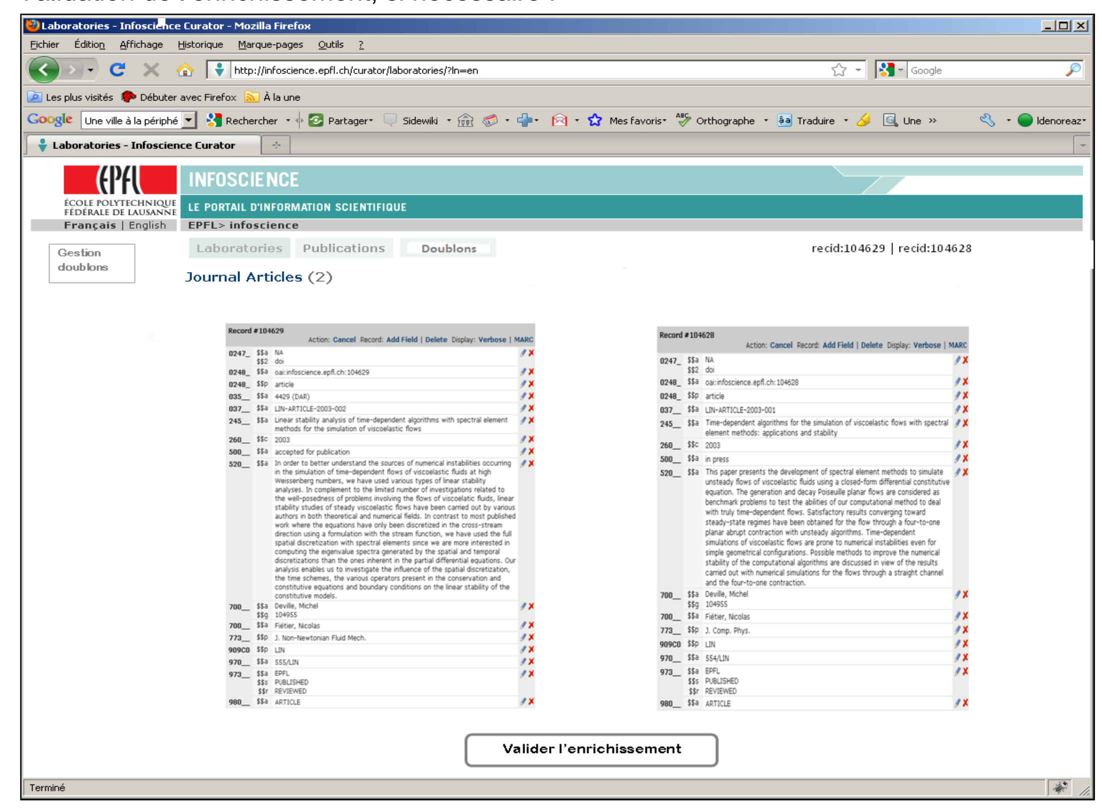


7. Evaluation/comparaison avec possibilité de passer en mode "Enrichir la notice" :

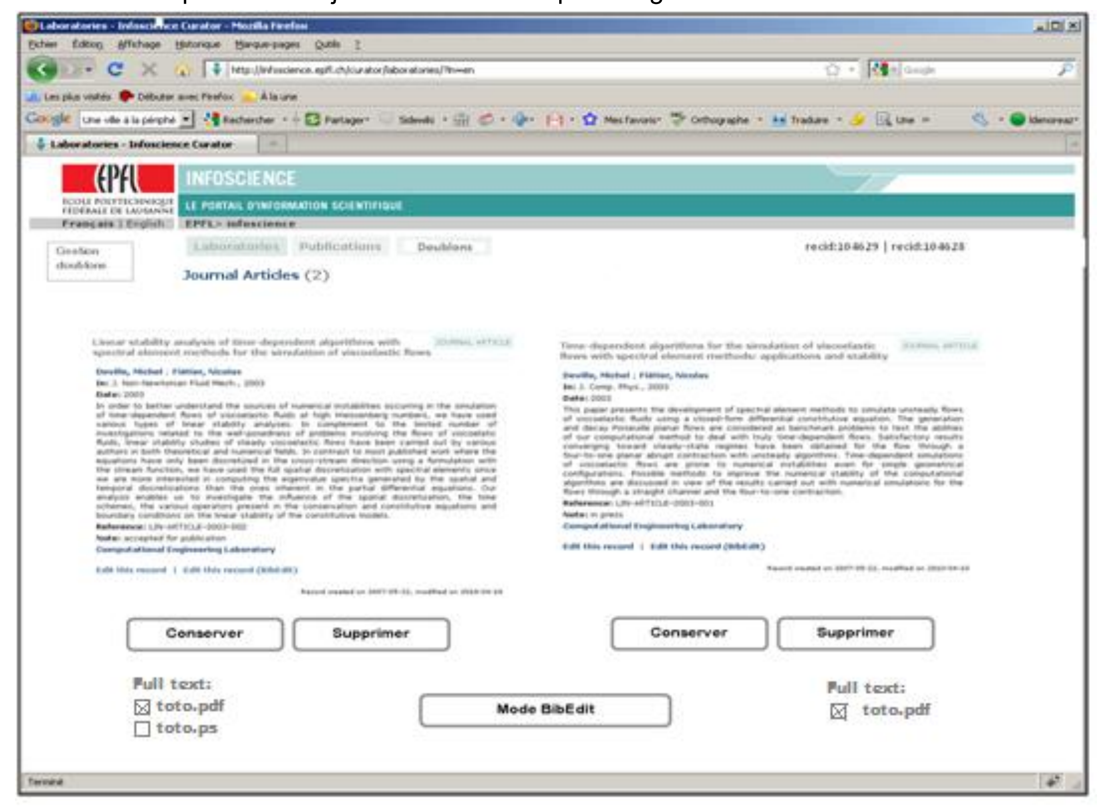


Pour enrichir il faut passer la souris sur la ligne du champ concerné. Une flèche apparaît, en cliquant dessus, le texte du champ sélectionné est automatiquement inséré dans l'autre notice. Si nécessaire, il est possible de supprimer une des lignes de la notice master ou forte en appuyant sur ✗.

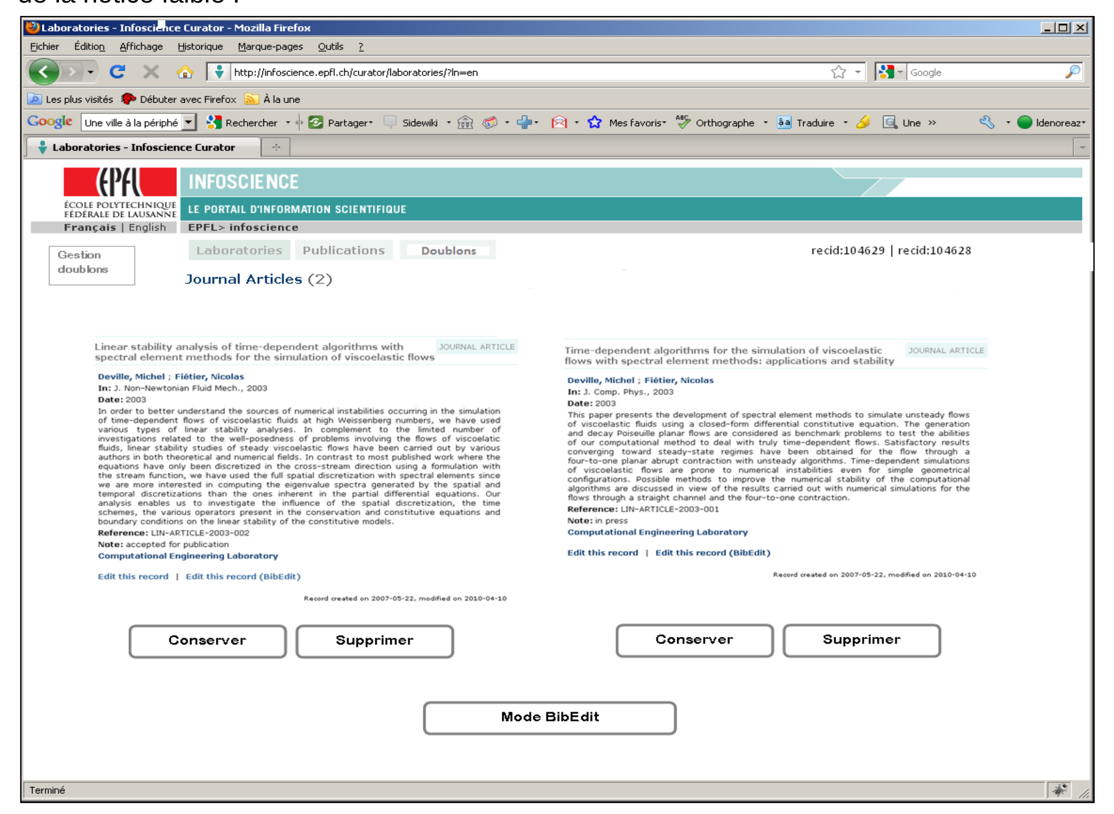
8. Validation de l'enrichissement, si nécessaire :



9. Affichage des éventuels fichiers en texte intégral liés à la notice et possibilité de sélectionner celui ou ceux que l'on veut joindre à la notice que l'on garde :



10. retour en mode HTML detailed après validation de l'enrichissement effectué, puis suppression de la notice faible :



7.6 Endnote X4

Il s'agit d'un logiciel qui permet de rechercher et de gérer des ressources bibliographiques. Etant donné qu'il intègre une fonctionnalité pour gérer des doublons, j'ai examiné cette dernière pour voir si je pouvais trouver des idées complémentaires pour les interfaces ci-dessus.

Voici quelques informations qui démontrent que les propositions énoncées ci-dessus pour les interfaces sont pertinentes:

Edit references when comparing duplicates—copy/paste and save changes in the side-by-side view.

- ⇨ Idéal d'avoir les notices en parallèle dans la même fenêtre

By default, a reference is considered a duplicate if the Author, Year, Title, and Reference Type match a reference already in the library.

- ⇨ Mêmes critères de bases pour comparer des notices

Le manuel d'utilisation Endnote (http://www.endnote.com/support/helpdocs/ENX4_GettingStartedGuide_Win.pdf) fournit de nombreux exemples d'interfaces qui semblent ergonomiques et conçues en vue du traitement de références. Par contre, il ne m'a pas été possible d'insérer des exemples d'interfaces tirées du manuel à cause du copyright.
Cette piste sera à creuser en continuant de travailler avec Grégory.

8 Bilan personnel

L'environnement de travail et l'ambiance ont été très agréables. J'ai bénéficié d'une grande liberté pour l'organisation de mon travail et d'une grande indépendance pour réaliser les tâches demandées. Tout en ayant si besoin le soutien de tous les collaborateurs de la Bibliothèque.

Pour varier un peu mon travail, j'ai pu officier au guichet d'accueil pendant une matinée. Cela m'a laissé voir une autre facette du métier et d'être en contact avec les usagers de la Bibliothèque.

Ce travail m'a surtout permis de me familiariser avec toute une série de notions abordées dans le cadre de mes études (périodiques, bibliométrie, facteur d'impact, peer-review, archives institutionnelles…) et de pouvoir creuser toute la problématique de mise à jour des informations dans un dépôt d'archives institutionnelles. J'ai acquis des connaissances tant au niveau informatique que bibliographique.

J'ai également compris l'importance pour une institution de la taille de l'EPFL d'avoir son dépôt d'archives institutionnelles. J'ai pu constater que les informations inclues dans Infoscience sont déjà très complètes grâce à la bonne collaboration entre l'équipe de conception et de développement du projet, les bibliothécaires scientifiques et les chercheurs qui alimentant cette base de données. Le sixième rang mondial du classement des dépôts en est bien la preuve.

9 Annexes

9.1 Cahier des charges

Stage

Laurence Denoréaz-Buclin

Bibliothèque de l'EPFL

19 juillet 2010 – 17 septembre 2010

Sous la direction de

David Aymonin

Avec le soutien de

Gregory Favre

Lionel Walter

Login : habituel epfl

Répertoire de travail : S:\Temporaire\Laurence

Enrichir la base de données des journaux

Objectifs

- Enrichir la base des journaux actuelle d'une information mentionnant si le journal subit un processus de peer-review ou non
- Voir si certaines autres informations pourraient être faciles à rajouter
  - Impact factor (par année)
  - Les noms abbrégés de certains journaux
  - Les informations sur le droit d'auteur (Sherpa/Romeo)
- Proposer des stratégies pour les mises à jour (une fois par an, plus souvent, est-ce qu'on peut charger uniquement les changements)

Outils

La base des journaux actuelle dans Infoscience (~10'000 journaux) : dump xml de Greg

xISSN : http://xissn.worldcat.org/xissnadmin/index.htm

Sherpa/Romeo : http://www.sherpa.ac.uk/romeo/

Ulrich's : http://www.ulrichsweb.com/

La base des journaux de la Bibliothèque de l'EPFL : http://library.epfl.ch/periodicals/ et la version backoffice sur "S:\SISB\Periodiques\Catalogues\db\periodicals.fp5" (personne de référence Georges Iffland)

Bimpe : http://bimpe.free.fr/bimpe_accueil.php

JCR (pour les impact factors) : http://isiknowledge.com/JCR

L'agence qui fournit les ISSN : http://www.issn.org/

(év) Web of Science : http://www.isiknowledge.com/

Scimago Journal Rank : http://www.scimagojr.com/

Rendez-vous avec Caroline Bosia qui gère les abonnements aux journaux à la Bibliothèque de l'EPFL

Etapes

Se familiariser avec les concepts de journaux scientifiques, d'ISSN, de peer-review, d'impact factor (wikipedia en anglais peut être un bon départ)

Apprendre à connaître les outils ci-dessus. Savoir quels sont leurs points forts, points faibles. Peut-on en extraire des listes ? Qu'a-t-on le droit de faire avec ces informations ? Sont-elles protégées par le copyright ?

Voir si d'autres systèmes existent (recherche documentaire). Que font les autres repository ? Des listes de repository peuvent être obtenues depuis

- Classement Webometrics : http://repositories.webometrics.info
- Directory of Open Access Repositories : http://www.opendoar.org/

Est-ce que des institutions telles le JISC (http://www.jisc.ac.uk/) proposent des choses dans ce domaine ?

Interface de déduplication

Objectifs

Dans infoscience, il y a environ 2000 documents qui sont à double (un exemple). Il faudrait avoir un moyen de les traiter pour pouvoir fusionner 2 documents se ressemblant en gardant les meilleures caractéristiques de chacun.

Il faudrait pouvoir signaler les faux doublons des vrais, ainsi que choisir quel champ l'on veut utiliser dans la notice finale. Garder à l'esprit que cette interface doit pouvoir aussi traiter les triplets, quadruplets…

Cette interface sera aussi utile lors de l'import de bases de données externes (notamment lors de la déduplication du Web of Science avec Infoscience).

Outils

- La liste des doublons d'Alain
- L'appel d'offre pour le curator : http://infoscience-wiki.epfl.ch/Curator?action=AttachFile&do=get&target=curator-CdC-v1.6.pdf page 12

Etapes

- Voir avec les « infoscience – librarians » qui ont déjà fait de la déduplication sur infoscience quels sont leurs besoins.
- Voir d'autres exemples d'outils faisant de la déduplication (Endnote, Jabref, … liste sous http://en.wikipedia.org/wiki/Comparison_of_reference_management_software). Peut-être s'inspirer d'autres outils de déduplications (outils de comparaison de code informatique, …)
- Faire une proposition d'interface
- Voir si elle serait fonctionnelle avec des exemples tirés de la liste des doublons d'Alain.

9.2 Glossaire

Ce glossaire contient les termes importants à connaître et maîtriser pour la réalisation du cahier des charges du stage Infoscience. Il répertorie sous un index classé par ordre alphabétique les notions concernées par le sujet des périodiques scientifiques. Les informations ont été directement recherchées sur Internet dans les moteurs de recherche Google et Exalead avec des équations de recherche simples.

Terme	Source	Définition (ou extrait de définition)
Bibliométrie (ou Infométrie)	http://www.adbs.fr/infometrie-17411.htm?RH=OUTILS_VOC	Traitements, par des méthodes statistiques mathématiques, et cartographiques pour la présentation des résultats, de l'information scientifique et technique (prise au sens large) jugée caractéristique de l'état de la science et de la technologie. Ces traitements, essentiellement descriptifs et classificatoires, s'appliquent à des informations quantitatives économiques, humaines et/ou bibliographiques. Dans ce dernier cas, on parle alors de **bibliométrie**. L'**infométrie** repose sur le principe d'une corrélation entre les éléments d'information traités et les concepts techniques ou scientifiques que ces éléments représentent. Elle permet d'alimenter des indicateurs quantitatifs (par comptage) ou relationnels (méthode des co-citations ou des mots associés, réseau d'acteurs) des activités de construction, de communication et d'usage des informations scientifiques et techniques, à partir desquels on peut, par exemple, établir la structure d'un domaine étudié, évaluer l'activité de recherche, déduire l'évolution d'une technologie ou d'un domaine d'activité, l'émergence d'un thème scientifique ou d'une technique, l'impact d'une publication, etc. Au sein de l'**infométrie**, la scientométrie traite la mesure des activités de recherche scientifique et technique.
	http://personal.rhul.ac.uk/uhtm/001/BibliometricsIndex.html	Bibliometrics is the quantitative analysis of bodies of literature (such as articles, books and patents) and their references: citations, and co-citations. Bibliometric analysis of literature allows the study of the foundations of a discipline and as a robust, quantitative approach augments the findings of more subjective literature reviews, In addition, when applied to patent data it also allows the investigation of firm and inventor networks by describing the linkages evident in citation to other individuals, firms and technologies. Bibliometrics can also be extended to graphically illustrate the most influential citations, how they are related, how strong their relationships are, and how far removed from, or central to, other groupings they are—in other words, the relationships inherent in the intellectual structure of a field or patent space. And co-citation studies can show us what topics, themes, and research methods are central, or peripheral, to a field, and how they may have changed over time.
	http://fr.wikipedia.org/wiki/Bibliom%C3%A9trie	La **bibliométrie** peut être définie comme « l'application des mathématiques et des méthodes statistiques aux livres, articles et autres moyens de communication » (Pritchard, 1969 [1]). Une définition plus contemporaine, associant bibliométrie, scientométrie d'outputs et infométrie, serait en termes d'analyse quantitative de l'activité et des réseaux scientifiques". Cette discipline, qui s'appuie sur l'analyse statistique des données et des réseaux, a un volet cognitif, en interaction avec les champs se donnant les sciences et les communautés scientifiques comme objet (économie de la connaissance, sociologie des sciences, épistémologie, histoire des sciences, etc.) et avec les sciences de l'information. Elle a un volet opérationnel, en liaison avec l'évaluation, le positionnement des acteurs et le management scientifique. Dans ce dernier cas, les études bibliométriques permettent de quantifier la communication d'un individu ou d'un groupe, non seulement en termes de volume, mais également de visibilité, d'influence, de partenariats, d'insertion dans les réseaux…
Métadonnée	http://fr.wikipedia.org/wi	Une **métadonnée** (mot composé du préfixe grec *meta*, indiquant l'auto-référence ; le mot signifie donc proprement

	ki/M%C3%A9tadonn%C3%A9e	« donnée de/à propos de donnée ») est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique).
	http://www.linux-france.org/prj/jargonf/M/meactadonneace.html	Information sur une information. Exemple type : le catalogue d'une bibliothèque est un ensemble de métadonnées. Les métadonnées sont cruciales pour le fonctionnement et la maintenance d'un datawarehouse. On devrait aussi en trouver sur le web, avec RDF et XML. Et on en trouve évidemment dans les système de fichiers, par exemple la date de dernière modification d'un fichier en est une.
	http://www.formastore.net/Metadonnee	Une métadonnée (du grec meta « après » et du latin data « informations ») est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique). Les métadonnées sont utilisées pour décrire et indexer les contenus, aussi bien dans les systèmes de gestion documentaire que dans les systèmes de gestion de contenus pédagogiques (CMS et LCMS). Les métadonnées des ressources pédagogiques remplissent différentes fonctions2 : • décrire et identifier (mots-clé, catégories, catalogue,...), • classer par collection ou catalogue • retrouver rapidement un contenu, en particulier grâce à des requêtes informatique ou au moyen d'un moteur de recherche. • gérer (type, auteur, date de mise à jour). • gérer les droits ; • permettre l'interopérabilité entre les plates-formes de télé-formation ou plates-formes pédagogiques3. La plupart des systèmes récents de métadonnées s'appuie sur des standards qui respectent les recommandations du « Dublin Core »
	http://www.educnet.education.fr/dossier/metadata/metadonnees	Une métadonnée est littéralement une donnée sur une donnée… Les métadonnées sont : - des informations qui décrivent un contenu : des objets concrets ou abstraits ; - des étiquettes qui permettent de retrouver des données.

ISSN	http://fr.wikipedia.org/wiki/International_Standard_Serial_Number	L'**ISSN** (International Standard Serial Number) est le numéro international qui permet d'identifier de manière unique une publication en série. Il concerne donc les journaux, les périodiques et les collections [par ex : Bibliothèque nationale suisse]. L'**ISSN** facilite la gestion informatique de ces publications (pour la gestion des abonnements, les citations de ces publications dans un texte, leur catalogage...), dans les bibliothèques et les services de documentation notamment. L'**ISSN** est normalisé par le texte ISO 3297 (ICS n° 01.140.20) L'**ISSN** est constitué des caractères « **ISSN** » suivis de deux groupes de quatre chiffres, ces groupes étant séparés par un tiret. Le dernier chiffre, situé en 8e position sert de clé de contrôle ; il peut parfois prendre la valeur « X », qui représente le nombre 10. Contrairement à l'ISBN, les chiffres de l'**ISSN** ne signifient rien en eux-mêmes. Ils sont attribués séquentiellement, indépendamment du pays d'origine, de la langue, etc. À chaque **ISSN** correspond un titre-clé ainsi qu'une période de publication, avec année de début et année de fin de publication. L'année de fin est conventionnellement fixée à « 9999 » lorsque la publication est toujours vivante (i.e. elle continue à paraître).
	http://www.issn.org/2-22635-What-is-an-ISSN.php	The **ISSN** is the standardized international code which allows the identification of any serial publication, including electronic serials, independently of its country of publication, of its language or alphabet, of its frequency, medium, etc... The **ISSN** is a numeric code which is used as an identifier: it has no signification in itself and does not contain in itself any information referring to the origin or contents of the publication. The **ISSN** takes the form of the acronym **ISSN** followed by two groups of four digits, separated by a hyphen. The eighth character is a control digit calculated according to a modulo 11 algorithm on the basis of the 7 preceding digits; this eighth control character may be an "X" if the result of the computing is equal to "10", in order to avoid any ambiguity. The **ISSN** is linked to a standardized form of the title of the identified serial, known as the "key title", which repeats the title of the publication, qualifying it with additional elements in order to distinguish it from other publications having identical titles. If the title of the publication changes in any significant way, a new **ISSN** must be assigned in order to correspond to this new form of title and avoid any confusion. A serial publication whose title is modified several times in the course of its existence will be assigned each time a new **ISSN**, thus allowing precise identification of each form of the title : in fact it is then considered that they are different publications even if there is a logical link between them.

		Contrary to other types of publications, the world of serial publications is particularly changeable and complex: the lifetime of a title may be extremely short; many publications may be part of a complex set of relationships, etc. These particularities themselves necessitated the introduction of the **ISSN** **ISSN** numbers are assigned by the **ISSN** national Centres coordinated in a network. All **ISSN** are accessible via the **ISSN** Register. The **ISSN** is not "just another administrative number". The **ISSN** should be as basic, a part of a serial, as the title.

		• As a standard numeric identification code, the **ISSN** is eminently suitable for computer use in fulfilling the need for file update and linkage, retrieval and transmittal of data. • As a human readable code, the **ISSN** also results in accurate citing of serials by scholars, researchers, information scientists and librarians. • In libraries, the **ISSN** is used for identifying titles, ordering and checking in, claiming serials, interlibrary-loan, union catalog reporting etc. • **ISSN** is a fundamental tool for efficient document delivery. **ISSN** provides a useful and economical method of communication between publishers and suppliers, making trade distribution systems faster and more efficient, in particular through the use of bar-coding and EDI (electronic data interchange). **ISSN** are assigned since the 1970s and universally accepted within the print publishing world as a means of identifying the serials. Changes in the editorial practices linked to the development of Internet led the **ISSN** to evolve and to extend its scope. The **ISSN** aims today at identifying and facilitating the access to the various versions and manifestations of the "continuing resources", which include the usual "serials" and the "ongoing integrating resources" available mostly online. Continuing resources are heterogeneous and their typology is still under realization. But they share the same characteristics with serials. They are "bibliographic resources issued over time with no predetermined conclusion...".
	http://www.nb.admin.ch/nb_professionnel/issn/01447/index.html?lang=fr	L'**ISSN** (International Standard Serial Number) est un numéro international normalisé qui identifie les publications en série de manière concise et sans équivoque. L'**ISSN** se compose de 8 chiffres dont le dernier est un caractère de contrôle qui peut aussi être la lettre X. L'**ISSN** facilite la gestion automatisée des publications en série dans les maisons d'édition, les agences, les librairies et les offices de distribution du domaine de la presse. Les institutions gérant des références bibliographiques utilisent les notices du registre **ISSN** pour leurs catalogues. L'**ISSN** permet aux étudiants et aux chercheurs de donner les références précises d'une publication en série. Accompagné de son titre-clé, l'**ISSN** est enregistré dans Helveticat, le catalogue de la Bibliothèque nationale suisse, ainsi que dans le Registre international de l'**ISSN** ; il permet donc la diffusion gratuite et mondiale de l'information sur la publication en question.
Norme ISO pour ISSN	**ISO 3297:2007**	L'ISO 3297:2007 définit et encourage l'emploi d'un code normalisé (**ISSN**), destiné à l'identification unique des publications en série et des autres ressources continues. Chaque Numéro international normalisé des publications en série (**ISSN**) est l'identifiant unique d'une publication en série ou d'une autre ressource continue sur un support déterminé. L'ISO 3297:2007 décrit aussi un mécanisme, l' «**ISSN** de liaison (ISSN-L)», qui permet de rassembler et de lier les versions sur des supports différents d'une même ressource continue. L'**ISSN** s'applique aux publications en série et aux autres ressources continues, qu'elles soient passées, présentes ou à paraître dans un avenir prévisible, quel que soit leur support de publication ou de production. Les monographies, les enregistrements sonores et visuels, les publications de musique imprimée, les œuvres audiovisuelles et les œuvres musicales ont leurs propres systèmes de numérotation et ne sont donc pas explicitement

		concernés par l'ISO 3297:2007. Ces documents peuvent porter un **ISSN** en complément de leurs propres numéros normalisés lorsqu'ils font partie d'une publication en série ou d'une ressource continue. Des recommandations plus détaillées sont fournies dans le manuel de l'**ISSN**, disponible auprès de l'autorité d'enregistrement de l'ISO 3297:2007.
Impact factor	http://en.wikipedia.org/wiki/Impact_factor	The **impact factor**, often abbreviated **IF**, is a measure reflecting the average number of citations to articles published in science and social science journals. It is frequently used as a *proxy (In statistics, a **proxy** variable is something that is probably not in itself of any great interest, but from which a variable of interest can be obtained.[1] In order for this to be the case, the proxy variable must have a close correlation, not necessarily linear or positive, with the inferred value)* for the relative importance of a journal within its field, with journals with higher impact factors deemed to be more important than those with lower ones. The **impact factor** was devised by Eugene Garfield, the founder of the Institute for Scientific Information (ISI), now part of Thomson Reuters. Impact factors are calculated yearly for those journals that are indexed in Thomson Reuter's *Journal Citation Reports*.
	http://thomsonreuters.com/products_services/science/academic/impact_factor/	The **impact factor** is a measure of the frequency with which the average article in a journal has been cited in a particular year or period. It is one of the evaluation tools provided by Thomson Reuters *Journal Citation Reports*® (*JCR*®). The annual *Journal Citation Reports* **impact factor** is a ratio between citations and recent citable items published: a journal's **impact factor** is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. The impact factor is a very useful tool for evaluation of journals, but it must be used discreetly. Considerations include the amount of review or other types of material published in a journal, variations between disciplines, and item-by-item impact. The journal's status in regard to coverage in the Thomson Reuters databases as well as the occurrence of a title change are also very important. In the next essay we will look at some examples of how to put tools for journal evaluation into use.
	http://fr.wikipedia.org/wiki/Facteur_d%27impact	Un **facteur d'impact**, également abrégé par les sigles **FI** ou **IF** (pour l'anglais *impact factor*) est une mesure de l'importance d'une revue scientifique. Il est calculé chaque année par l'Institute for Scientific Information (ISI) pour chaque revue suivie. Il est publié dans le *Journal Citation Reports*. Les facteurs d'impact ont une grande influence dans le domaine de la publication scientifique car ils servent souvent de critères d'évaluation *quantitatifs*. Le facteur d'impact est calculé sur une période de trois ans. Par exemple, le facteur d'impact de 2003 pour une revue X est calculé ainsi . *A* = le nombre de fois que des articles publiés dans X durant la période 2001-2002 sont cités dans l'ensemble des revues indexées durant l'année 2003. *B* = le nombre d'articles, résumés étendus ou notes publiés dans X dans la période 2001-2002. Le facteur d'impact de la revue X pour 2003 sera *A*/*B* L'ISI exclut certains types d'articles comme les erratas dans le calcul du dénominateur B. Les nouvelles revues, indexées depuis leur premier numéro, recevront un facteur d'impact après une période de deux ans. Comme exemple, les deux revues scientifiques considérées les plus prestigieuses sont *Nature* et *Science*. Toutes les deux ont des facteurs d'impact autour de 30. Certaines revues bio-médicales ont des valeurs plus importantes. Nombres

		de revues de domaines spécialisés de haut niveau ont des facteurs d'impact inférieurs à 5.
Peer review	http://en.wikipedia.org/wiki/Peer_review	**Peer review** is a generic term that is used to describe a process of self-regulation by a profession or a process of evaluation involving qualified individuals with the related field. **Peer review** methods are employed to maintain standards, improve performance, and provide credibility. Scholarly **peer review** (also known as **refereeing**) is the process of subjecting an author's scholarly work, research, or ideas to the scrutiny of others who are experts in the same field, before a paper describing this work is published in a journal. **Peer review** requires a community of experts in a given (and often narrowly defined) field, who are qualified and able to perform impartial review. Impartial review, especially of work in less narrowly defined or inter-disciplinary fields, may be difficult to accomplish; and the significance (good or bad) of an idea may never be widely appreciated among its contemporaries. Although generally considered essential to academic quality, and used in most important scientific publications, **peer review** has been criticized as ineffective, slow, and misunderstood (see anonymous peer review and open peer review). Pragmatically, **peer review** refers to the work done during the screening of submitted manuscripts and funding applications. This process encourages authors to meet the accepted standards of their discipline and prevents the dissemination of irrelevant findings, unwarranted claims, unacceptable interpretations, and personal views. Publications that have not undergone **peer review** are likely to be regarded with suspicion by scholars and professionals... **Procedure** In the case of proposed publications, an editor sends advance copies of an author's work or ideas to researchers or scholars who are experts in the field (known as "referees" or "reviewers"), nowadays normally by e-mail or through a web-based manuscript processing system. Usually, there are two or three referees for a given article...
	http://fr.wikipedia.org/wiki/%C3%89valuation_par_les_pairs	Dans les disciplines scientifiques, l'évaluation par les pairs désigne l'activité collective des chercheurs qui jaugent de façon critique les travaux d'autres chercheurs (leurs « pairs »). Ces évaluations peuvent porter sur une recherche précise soumise pour publication dans une revue scientifique ou destinée à être présentée à une conférence mais elles peuvent aussi couvrir l'ensemble des travaux du chercheur ou du groupe de chercheurs évalués, notamment lors du recrutement d'un candidat à un poste ou lors de l'évaluation de projets de recherche par des institutions publiques (comme le CNRS) ou privées (comme une fondation). Pour les revues scientifiques, l'évaluation par les pairs est menée par des *comités de lecture* qui décident si le compte-rendu d'un travail de recherche soumis pour publication est acceptable ou non.
	http://www.parliament.uk/documents/post/pn182.pdf	**Peer review** is a system whereby research – or a research proposal - is scrutinised by (largely unpaid) independent experts (peers). In general, the process serves a technical (ensuring that the science is sound) and a subjective function (is the science interesting, important and/or groundbreaking?). The flowchart below gives a brief overview of how the process works to select science for funding and publication, although in practice, there is considerable variation in **peer review** processes between funding bodies and journals.
Journal scientifique ou revue	http://fr.wikipedia.org/wiki/Revue_scientifique	Une **revue scientifique** est une revue académique destinée à publier des travaux de recherche originaux et à animer le débat des spécialistes. Les revues scientifiques, dans leur grande majorité, sont spécialisées dans un aspect particulier d'un domaine : sciences, technologie, médecine (STM), sciences humaines et sociales (SHS). Les articles publiés sont

scientifique		examinés auparavant par un comité de lecture composé de chercheurs de la même spécialité. Une revue scientifique ne doit pas être confondue avec un article de revue, qu'elle peut publier. La publication d'un travail de recherche dans une revue scientifique permet de le confronter à la communauté scientifique, et d'archiver les résultats pour référence ultérieure. Dans le cadre de l'évaluation des performances de recherche, on tente de comptabiliser le nombre de publications, l'impact des revues scientifiques dans lesquelles elles sont publiées (c'est-à-dire les lectorats potentiels des articles) et le nombre de publications ultérieures citant chacune d'elles. Cette technique d'évaluation, parfois décriée pour ses biais, est appelée bibliométrie. Ces revues peuvent être l'émanation de sociétés savantes (SHS) ou d'académies des sciences, par exemple, mais peuvent aussi avoir été fondées indépendamment. Des éditeurs de presse se spécialisent dans leur édition, impression et distribution.

	http://en.wikipedia.org/wiki/Scientific_journal	In academic publishing, a **scientific journal** is a periodical publication intended to further the progress of science, usually by reporting new research. There are thousands of scientific journals in publication, and many more have been published at various points in the past (see list of scientific journals). Most journals are highly specialized, although some of the oldest journals such as *Nature* publish articles and scientific papers across a wide range of scientific fields. Scientific journals contain articles that have been peer reviewed, in an attempt to ensure that articles meet the journal's standards of quality, and scientific validity. Although scientific journals are superficially similar to professional magazines, they are actually quite different. Issues of a scientific journal are rarely read casually, as one would read a magazine. The publication of the results of research is an essential part of the scientific method. If they are describing experiments or calculations, they must supply enough details that an independent researcher could repeat the experiment or calculation to verify the results. Each such journal article becomes part of the permanent scientific record. The history of scientific journals dates from 1665, when the French *Journal des sçavans* and the English *Philosophical Transactions of the Royal Society* first began systematically publishing research results. Over a thousand, mostly ephemeral, were founded in the 18th century, and the number has increased rapidly after that.[1] Articles in scientific journals can be used in research and higher education. Some classes are partially devoted to the explication of classic articles, and seminar classes can consist of the presentation by each student of a classic or current paper. In a scientific research group or academic department it is usual for the content of current scientific journals to be discussed in journal clubs. The standards that a journal uses to determine publication can vary widely. Some journals, such as *Nature*, *Science*, *PNAS*, and *Physical Review Letters*, have a reputation of publishing articles that mark a fundamental breakthrough in their respective fields. In many fields, an informal hierarchy of scientific journals exists; the most prestigious journal in a field tends to be the most selective in terms of the articles it will select for publication, and will also have the highest **impact factor**. It is also common for journals to have a regional focus, specializing in publishing papers from a particular country or other geographic region, like *African Invertebrates*. Articles tend to be highly technical, representing the latest theoretical research and experimental results in the field of science covered by the journal. They are often incomprehensible to anyone except for researchers in the field and advanced students. In some subjects this is inevitable given the nature of the content. Usually, rigorous rules of scientific writing are enforced by the editors, however these rules may vary from journal to journal, especially between journals from different publishers.
Web service	http://fr.wikipedia.org/wiki/Service_Web	Un service web (ou service de la toile[1]) est un programme informatique permettant la communication et l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués. Il s'agit donc d'un ensemble de fonctionnalités exposées sur internet ou sur un intranet, par et pour des applications ou machines, sans intervention humaine, et en temps réel. Voir aussi : http://www.w3.org/standards/webofservices/

	http://www.commentcamarche.net/contents/web-services/web-service.php3	Les **services web** (en anglais *web services*) représentent un mécanisme de communication entre applications distantes à travers le réseau internet indépendant de tout langage de programmation et de toute plate-forme d'exécution : • utilisant le protocole HTTP comme moyen de transport. Ainsi, les communications s'effectuent sur un support universel, maîtrisé et généralement non filtré par les pare-feux ; • employant une syntaxe basée sur la notation XML pour décrire les appels de fonctions distantes et les données échangées ; • organisant les mécanismes d'appel et de réponse. Grâce aux services web, les applications peuvent être vues comme un ensemble de services métiers, structurés et correctement décrits, dialoguant selon un standard international plutôt qu'un ensemble d'objets et de méthodes entremêlés. Le premier bénéfice de ce découpage est la facilité de maintenance de l'application, ainsi que l'interopérabilité permettant de modifier facilement un composant (un service) pour le remplacer par un autre, éventuellement développé par un tiers. Qui plus est, les services web permettent de réduire la complexité d'une application car le développeur peut se focaliser sur un service, indépendamment du reste de l'application. Les services web facilitent non seulement les échanges entre les applications de l'entreprise mais surtout permettent une ouverture vers les autres entreprises. Les premiers fournisseurs de services web sont ainsi les fournisseurs de services en ligne (météo, bourse, planification d'itinéraire, pages jaunes, etc.), mettant à disposition des développeurs des API (Application Programmable Interface) payantes ou non, permettant d'intégrer leur service au sein d'applications tierces. Voir aussi : • http://www.softeam.fr/technologies_web_services.php • http://www.webopedia.com/TERM/W/web_services.html
API	http://fr.wikipedia.org/wiki/Interface_de_programmation	Une **interface de programmation** (*Application Programming Interface* ou *API*) est une interface fournie par un programme informatique. Elle permet l'interaction des programmes les uns avec les autres de manière analogue à une interface homme-machine rend possible l'interaction entre un homme et une machine. Du point de vue technique une API est un ensemble de fonctions, procédures ou classes mises à disposition par une bibliothèque logicielle, un système d'exploitation ou un service. La connaissance des API est indispensable à l'interopérabilité entre les composants logiciels**.**
	http://en.wikipedia.org/wiki/Application_programming_interface	An **application programming interface** (**API**) is an interface implemented by a software program which enables it to interact with other software. It facilitates interaction between different software programs similar to the way the user interface facilitates interaction between humans and computers. An API is implemented by applications, libraries, and operating systems to determine their vocabularies and calling conventions, and is used to access their services. It may include specifications for routines, data structures, object classes, and protocols used to communicate between the consumer and the implementer of the API.

DOI	http://fr.wikipedia.org/wiki/Digital_Object_Identifier	***Digital Object Identifier*** (**DOI**, littéralement « identifiant d'objet numérique ») est un mécanisme d'identification de ressources numérisées, comme un film, un rapport, des articles scientifiques, etc. C'est une alternative aux URI. Depuis février 2010, l'Institut de l'information scientifique et technique (INIST, du CNRS), est doté d'un statut « agence DOI », faisant partie du consortium DataCite. Les DOI permettent notamment de faciliter l'utilisation des bases de données bibliographiques ainsi que les logiciels de gestion bibliographique.
		A **digital object identifier** (**DOI**) is a character string used to uniquely identify an electronic document or other object. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document is permanent, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the metadata for the DOI to link to the new URL. However, unlike URLs, the DOI system is not open to all comers; only organizations that can meet the contractual obligations of the DOI system and that are willing to pay to become a member of the system can assign DOIs. The DOI system is implemented through a federation of registration agencies coordinated by the International DOI Foundation, which developed and controls the system. The DOI system has been developed and implemented in a range of publishing applications since 2000; by late 2009 approximately 43 million DOI names had been assigned by some 4,000 organizations. A DOI name takes the form of a character string divided into two parts: a prefix and a suffix. The prefix identifies the registrant of the name, and the suffix is chosen by the registrant and identifies the specific object associated with that DOI. Most legal Unicode characters are allowed in these strings, which are interpreted in a case-insensitive manner. Major applications of the DOI system currently include: • persistent citations in scholarly materials (journal articles, books, etc.) through CrossRef, a consortium of around 3,000 publishers; • scientific data sets through DataCite, a consortium of leading research libraries, technical information providers, and scientific data centers; • European Union official publications through the EU publications office. Et aussi : • http://www.doi.ethz.ch/index_f.html • http://scitation.aip.org/jhtml/doi.jsp (éditeurs et utilisation des DOI) • pour trouver à quoi correspond un DOI: http://dx.doi.org/

9.3 Email aux autres « repository »

Version française

Bonjour,

Dans le cadre de mes études de spécialiste en information documentaire, je suis en train d'effectuer un stage professionnel à la fin de ma deuxième année d'études.

Ce stage qui se déroule à la bibliothèque du Rolex Learning Center de l'Ecole polytechnique fédérale de Lausanne (http://www.epfl.ch/index.fr.html) a pour mandat d'enrichir la base de journaux/périodiques utilisés pour le portail scientifique de l'EPFL : Infoscience (http://infoscience.epfl.ch/).

Pour me permettre de répondre aux exigences du cahier des charges, j'aurai besoin de connaître un certain nombre d'informations et je vous remercie d'avoir l'obligeance de m'indiquer de quelle manière votre institution s'approvisionne concernant les points suivants :

1. Le facteur d'impact
2. Le processus peer-review
3. Les numéros ISSN
4. Les noms abrégés de journaux
5. Le droit d'auteur

Est-ce que vous utilisez des :

- services web
- applications gratuites ou payantes
- abonnements à des services en ligne
- autres, et si oui lesquels ?

Actuellement, l'EPFL alimente sa base de journaux et les informations y-relatives via les outils suivants :

Les numéros ISSSN : xISSN : http://xissn.worldcat.org/xissnadmin/index.htm
Le droit d'auteur: Sherpa/Romeo : http://www.sherpa.ac.uk/romeo/

Je vous remercie pour le temps consacré à éclaircir mon besoin d'informations et je reste à votre disposition pour envisager un partage des informations entre votre institution et le dépôt institutionnel de l'Ecole polytechnique fédérale de Lausanne.

Avec mes meilleures salutations,
Laurence Denoréaz

Version anglaise

Hello,

As part of my bachelor in science and information management, I am running a project sponsored by the EPFL Library in the Rolex Learning Center. The objective is to expand the collections of journals and reviews offered by the scientific portal of EPFL, called InfoScience (http://infoscience.epfl.ch). In this context, I'd be interested in learning more about how your organization is addressing the following topics:

- Impact factor
- Peer review process
- ISSN numbers
- Journal names abbreviations
- Copyrights

Also, do you make use of:

- Web services
- Specific applications (free or licensed)
- On-line services subscriptions

For your information, the EPFL Library is currently feeding its collection from the following sources:

- ISSN numbers from xISSN (http://xissn.worldcat.org/xissnadmin/index.htm)
- Copyrights from Sherpa/Romeo (http://www.sherpa.ac.uk/romeo)

I'd like to thank you in advance for the information you will provide me, and I'm available to discuss any sharing of information between your organization and the institutional depot of EPFL.

Kind regards,

Laurence Denoreaz

10 Liste non exhaustive de liens complémentaires

- **DOIs :**

  http://www.crossref.org/

- **Droit d'auteur :**

  http://www.abf.asso.fr/pages/interieur-contenu.php?categorieTOP=6&categorie=46&id_contenu=41

  http://www.abf.asso.fr/pages/interieur-contenu.php?categorieTOP=6&categorie=46&id_contenu=40

  http://www.abf.asso.fr/pages/interieur-contenu.php?categorieTOP=6&categorie=62&id_contenu=69

  http://www.bis.info/fr/contenu-themes-professionnels/droit-dauteur

  http://www.droitauteur.levillage.org/

  http://www.copyright.com/viewPage.do?pageCode=bu1

- **Facteur d'impact et bibliométrie :**

  http://cybergeo.revues.org/index15463.html

  http://cybergeo.revues.org/index22862.html

  http://www.histcite.com/

  http://hlwiki.slais.ubc.ca/index.php/Impact_factor#E-index

  http://personal.rhul.ac.uk/uhtm/001/BibliometricTechniques.html

  http://www4.rhbnc.ac.uk/~uhtm001/bibexcel-primer.pdf

- **Journaux scientifiques :**

  http://www.doaj.org/

  http://www.scicentral.com/Y-journa.html

  http://www.emeraldinsight.com/librarians/index.htm

  http://www.emeraldinsight.com/help/librarian/index.htm

  http://www.emeraldinsight.com/help/pdf/step_by_step_FR.pdf

  http://www.scientificjournals.org/

11 Bibliographie

11.1 Lectures complémentaires

Sylvie Godel, Isabelle de Kaenel, Pablo Iriarte. Web services for bibliometrics. In : EAHIL2010 [en ligne]. http://archivesic.ccsd.cnrs.fr/sic_00540289/ (consulté le 21 juillet 2010)

Péter Jacsó, (2010) "Comparison of journal impact rankings in the SCImago Journal & Country Rank and the Journal Citation Reports databases", Online Information Review, Vol. 34 Iss: 4, pp.642 – 657 (consulté le 28 juillet 2010)

Yves-François Le Coadic. Défense et illustration de la bibliométrie. In : Bulletin des Bibliothèques de France [en ligne]. http://bbf.enssib.fr/consulter/bbf-2010-04-0048-014 (consulté le 20 août 2010)

11.2 Sites consultés durant le stage

Outils :

BIMPE. Base d'Information Mutualiste sur les Périodiques Electroniques. [en ligne]. http://bimpe.free.fr/bimpe_accueil.php (consulté en juillet, août et septembre 2010)

ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE. BIBLIOTHEQUE DE L'EPFL. *Site de la Bibliothèque* [en ligne]. http://library.epfl.ch/periodicals/ (consulté en juillet, août et septembre 2010)

ELSEVIER. SCOPUS [en ligne]. http://www.scopus.com/home.url (consulté en juillet, août et septembre 2010)

ISI Web of Knowledge. Journal Citation Report [en ligne]. http://isiknowledge.com/JCR (consulté en juillet, août et septembre 2010)

ISI Web of Knowledge [en ligne]. http://www.isiknowledge.com/ (consulté en juillet, août et septembre 2010)

ISSN International Center. ISSN.org [en ligne].http://www.issn.org/ (consulté en juillet, août et septembre 2010)

JISC [en ligne]. http://www.jisc.ac.uk/ (consulté en juillet, août et septembre 2010)

SHERPA/ROMEO. Publisher copyright policies & self-archiving [en ligne]. http://www.sherpa.ac.uk/romeo/ (consulté en juillet et août 2010)

SJR. SCImago Journal & Country Rank [en ligne]. http://www.scimagojr.com/ (consulté en juillet, août et septembre 2010)

ULRICHSWEB. Global serials directory [en ligne]. http://www.ulrichsweb.com/ (consulté en juillet, août et septembre 2010)

ULRICH'S SERIALS ANALYSIS SYSTEM [en ligne]. http://www.ulrichsweb.com/ulrichsweb/analysis/default.asp?navPage=4& (consulté en juillet, août et septembre 2010)

WORLDCAT. xISSN Web service [en ligne]. http://xissn.worldcat.org/xissnadmin/index.htm (consulté en juillet et août 2010).

Outil de développement :

CDS SOFTWARE CONSORTIUM. CDS INVENIO [en ligne]. http://cdsware.cern.ch/invenio/ (consulté en juillet 2010)

Droit d’auteur :

Loi fédérale sur le droit d’auteur et les droits voisins. RS 231.1 (consulté en septembre 2010)

Archives institutionnelles :

ARCHIMER. Archive institutionnelle de l’IFREMER [en ligne]. http://archimer.ifremer.fr/ (consulté en août 2010)

HAL. Hyper Articles en Ligne [en ligne]. (consulté en août 2010)

OATOA. Open Archive Toulouse Archive Ouverte.[en ligne]. http://oatao.univ-toulouse.fr/ (consulté en août 2010)

ORBI. Open Repository and Bibliography [en ligne]. http://orbi.ulg.ac.be/ (consulté en août 2010)

Ranking Web of World repositories [en ligne]. http://repositories.webometrics.info/(consulté en juillet et août 2010)

SERVAL. Serveur académique lausannois [en ligne]. UNIL-SERVAL (http://www.unil.ch/serval/)(consulté en juillet 2010)

UNIVERSITE DE ST-Gall. Plateforme de recherche Alexandria. [en ligne]. http://www.alexandria.unisg.ch/ (consulté en août 2010)